STUDY ON THE RELATIONSHIP BETWEEN LEFT-TURN TRAFFIC OPERATIONS AND SAFETY AT SIGNALIZED INTERSECTIONS

A Thesis

by

SUNGHOON LEE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

August 2007

Major Subject: Civil Engineering

STUDY ON THE RELATIONSHIP BETWEEN LEFT-TURN TRAFFIC OPERATIONS AND SAFETY AT SIGNALIZED INTERSECTIONS

A Thesis

by

SUNGHOON LEE

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Approved by:

Chair of Committee,	Yunlong Zhang
Committee Members,	Dominique Lord
	Tom Wehrly
Head of Department,	Anthony Cahill

August 2007

Major Subject: Civil Engineering

ABSTRACT

Study on the Relationship Between Left-Turn Traffic Operation and Safety at Signalized Intersections. (August 2007)

Sunghoon Lee, B.S., Hong-ik University, Seoul Korea

Chair of Advisory Committee: Dr. Yunlong Zhang

Intersections are the most complex locations in a traffic system and are likely to have a higher crash count than any other location in the system. Intersection safety is related to traffic operations, such as traffic signal and approaching volume. The objective of this study is to determine the contributing factor for left-turn crashes at signalized intersections by a statistical modeling process and to develop crash prediction models. Potential contributing factors representing the characteristic of a left-turn operation were identified and considered for inclusion in crash prediction models. HCS (Highway Capacity Software) 2000 was utilized for computing some traffic indicators such as volume to capacity ratio for potential inclusion in the models. Three years of crash data were collected in the College Station area. The Signal timing and Volume data were obtained from public works in College Station. The volume data was sorted into three time periods and signal timing data were obtained for three different time periods: AM, noon, and PM.

The division of time periods results from timing plans being changed for different periods. Relationship between crash count and each factor was explored to identify whether the factor has the potential to influence the crash count. Afterwards, the prediction models were developed using the negative binomial structure because of many zero samples. Akaike Information Criteria was used for selecting the model having the best fit. Wald tables provided that variables have significance in affecting the left-turn crash count. Left-turn type, sequence, volume, control delay, and post speed limit were identified as significant factors impacting left-turn crash count at a signalized intersection.

ACKNOWLEDGMENTS

First of all, I want to give thanks to God who is governing the world. He is the reason I am where I am today. He has always encouraged me, inspired me, and given me the opportunity and capabilities of studying for this thesis process and completing my degree. Secondly, I want to thank my advisor Dr. Zhang. He has always tried to help me out and provide guidance to complete my thesis. I also would like to thank Dr. Lord and Dr. Wehrly for providing the help in the areas of safety analysis and statistical modeling process and for serving on my committee. Third of all, I appreciate my parents. They gave birth to me and made me the person I am, and have also supported me in many ways. They have always been my most important mentors and advisors. Finally, I appreciate Sarah for editing my thesis. Without her help, I could not have made it.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Signalized intersections are one of the most complex spaces on transportation network. There are 12 vehicle movements at a typical four-legged intersection, causing many conflicts that may lead to crashes (1). Although the intersection may be hazardous to vehicles due to conflicts and turning movements, traffic control devices such as signals reduce potential crashes by eliminating conflicts through allocating right of way to movements at a given time. Signalized intersections are particularly unique and they can be intelligent elements in a traffic system in being changeable depending on the traffic's situations. This is how signals differ from other traffic control devices such as signs and markings. In addition, a signal can be utilized in other applications such as ramp metering. Traditionally, signal studies have mainly focused on the operational aspects such as increasing capacity, optimizing signal timing, alleviating traffic congestion, and improving travel time. The operational benefits of traffic signal can be easily estimated by software programs and manual procedures. Safety effects related to signal operations, however, can not be presented as clearly for different traffic conditions.

In past studies, left-turns at an intersection have raised many significant safety questions. Nonetheless, there have been only a few left-turn crash models developed and those models only included very few operational parameters aside from volumes. This study mainly focuses on left-turn crashes and traffic factors associated with left-turns such as LT capacity, LT v/c ratio, control delay, intersecting volume, and signal timing parameters. While previous safety models considered volume as the only explanatory variable, this study considers other traffic factors along with the traffic volume based on the assumption that volume cannot fully represent the nature of accidents (2). Those factors were employed in an attempt to improve the safety model. The objective of this study is to develop a feasible model to estimate left-turn crashes as a function of factors influence the operations at traffic signals.

This thesis follows the style of Transportation Research Record.

Problem Statement

Many researchers have attempted to identify significant factors relating to crashes at intersections through statistical models. Based on data obtained from the College Station Police Department from recent years, left-turn related crashes account for over 40% of intersection crashes in College Station (3). For that reason, understanding left-turn crash causality could lead to treatments that would significantly reduce overall crash counts at intersections. The possible causes of crashes at intersections have been identified in many studies. (e.g. E. Hauer, D. Lord, Lyon C., Bauer K.) However, little work had been focused on left-turn crashes and developing model for estimating and predicting left-turn crashes at signalized intersections. This study used many traffic elements associated with the left-turn. From the modeling procedure, several models have been developed for left-turn crashes. The best model can be chosen using various goodness-of-fit tests. The model selection criteria do not always depend on the basis of a better fit, but follow a logical process. As mentioned before, a good understanding of left-turn crash causality will be very helpful for engineers and decision makers in identifying potential safety problems and developing countermeasures at signalized intersections. This research was designed to determine significant factor in affecting left-turn crashes at signalized intersections.

Research Objective

The goal of this study is to identify significant factors related to the safety of left-turns at signalized intersections. Using the identified factors, safety models will be developed with the aid of statistical methods. In the modeling process, each coefficient and its corresponding effects will be estimated. The detailed objectives are presented as follows:

1. Identifying the potential traffic elements through exploring the collected data;
2. Presenting the relationship between potential factors and left-turn crashes;
3. Finding the factors impacting left-turn crashes;
4. Developing the prediction model for left-turn crashes; and
5. Estimating the impact of factors the left-turn crash used in the model.

The crash data will be analyzed using separate time periods such as morning peak and evening peak because same-intersection crash characteristics may vary with flow fluctuations throughout the day and the signal timing for different time periods may also be different. The data will be further disaggregated by left-turn phasing sequence or type in order to model the effects of different LT treatments. In order to accomplish the goal of this research, the following tasks will be performed:

- Conduct a thorough review of related previous studies;
- Select the target study area;
- Obtain data for the selected sites;
- Reduce and analyze data;
- Develop crash models;
- Present findings; and
- Propose recommendations for the potential inclusion into the design process.

Thesis Organization

This thesis is composed of six chapters. Chapter I addresses the background. It also describes the problem statement and research objectives. The second chapter of the thesis provides a review of previous research and state-of-the art on intersection operation and safety studies. The third chapter described preliminary analysis of the collected data in order to identify potential factors that affect left-turn crash counts. The fourth chapter explains the methodology employed for this study. The chapter also illustrates statistical approach and modeling process used the in traffic safety area. The fifth chapter addresses the results from the exploratory data analysis and estimation of expected left-turn crash counts based on a developed predictive model. Lastly, the sixth chapter states the executive summary of this research. This summary also includes limitations and suggestions for future work.

CHAPTER II

LITERATURE REVIEW

This chapter briefly describes the basics of signal operations, particularly the left-turn operations. It will then focus on summarizing the past research in the area of the left-turn safety study and crash prediction models at the signalized intersection. Those studies reviewed mainly present the statistical approaches for the safety model development and estimation.

Signalized Intersection and Left-Turn Operations

A traffic signal is a critical form of intersection control. It can considerably alleviate safety problems of an intersection on account of assigning the right-of-way to specific movements at a given time. No other form of control can do as much as signals do. Drivers can avoid some of the most severe conflicts as long as they obey the signal. Traffic control can not remove all types of conflicts, from trivial conflict to the most severe crashes. Nevertheless obeying the traffic signal is the best way to stay away from a traffic accident (1). The following sections will explain the definitions, functions , and implications of those elements related to traffic operations that may affect the safety at signalized intersections.

Left-turn Sequence and Operations

A left-turn is the most difficult and complex procedure to deal with at a signalized intersection (1). Several different elements are involved in left-turn operations. Different designs can be considered when dealing with a left-turn at an intersection. A left-turn movement can be given their own lane (exclusive a left-turn lane) or share the lane with the through movements in a shared lane. A traffic signal can have different left-turn phasing strategies: permitted left-turn, protected left-turn and permitted plus protected left-turn depending on traffic situations such as volume, service capacity, the intersection geometry design. A left-turn movement can use a shared lane or an exclusive lane in terms of how the lane is assigned to a left-turn vehicle. The left-turn phasing selection depends on left-turn volume, the opposing through volume, and local practice. Lastly, the left-turn movement consumes more time for going through the intersection than the through movement. Consequently, the left-turn movement volume needs to be adjusted by a factor called the left-turn equivalence (1). If the left-turn is protected with an exclusive lane and the two left-turn flows have a considerable difference, inefficiency is likely to exist in traffic operation if concurrent left-turn green is used. For example, if one of two approaches has a large volume, then left-turn movements will be provided a long green duration on both directions. However, the approach with low volume does not need as much green time as one with a larger volume. Thus, some green time for the approach with lower demand will be wasted. To enhance efficiency, the two left-turn phases should be timed separately. The example of non-concurrent left-turn signal phasing scheme include lead-lag left-turns (Figure 1) and dual lead left-turns with overlapping (Figure 2). Different left-turn configurations and treatments affect the operations and safety of vehicles at signalized intersections.

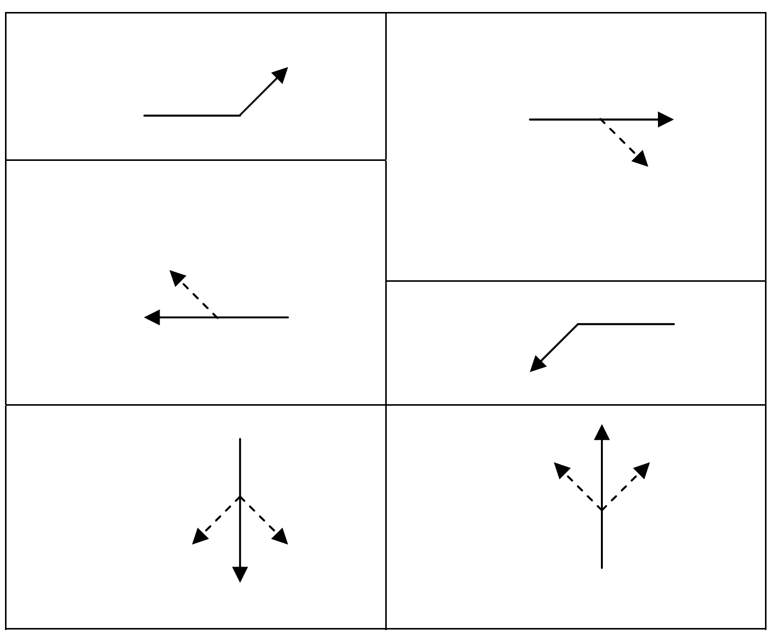

FIGURE 1 Lead/lag green phasing

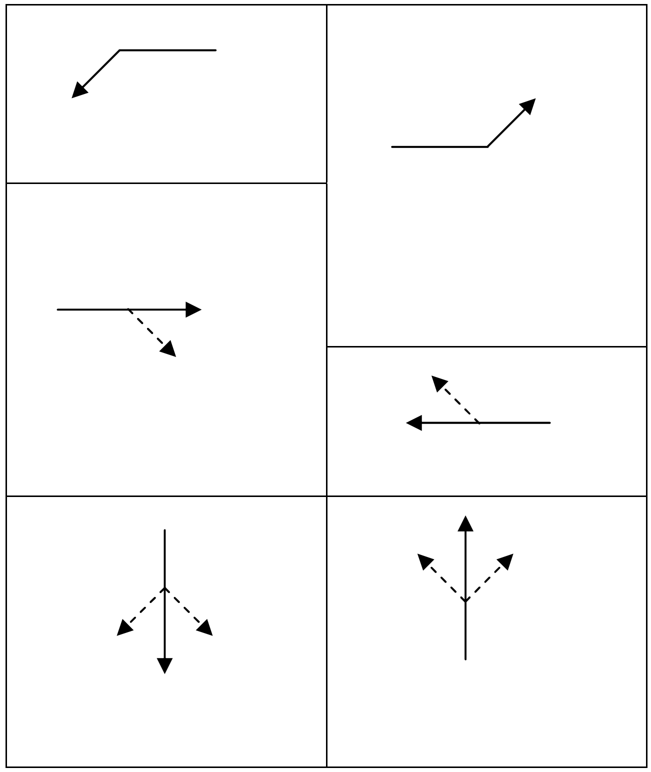

FIGURE 2 Lead green with overlapping phasing

Control Delay

Control delay is experienced by vehicles at signalized intersections. It is caused by a control device, either a traffic signal or a stop-sign. The control delay time includes the stopped time and the time lost for deceleration or acceleration. The Highway Capacity Manual (2000) defines the control delay as follows (4):

> *The control delay is the portion of the total delay for a vehicle approaching and entering a signalized intersection that is attributable to a traffic signal operation.*

Control delay includes the delays of initial deceleration, move-up time in the queue, stops, and reacceleration.

According to the HCM, delay is calculated by the following equation:

$$d = d_1(PF) + d_2 + d_3 \tag{1}$$

$$d_1 = \frac{0.5C\left(1 - \frac{g}{C}\right)^2}{1 - \left[\min(1, X)\frac{g}{C}\right]} \tag{2}$$

$$d_2 = 900T\left[(X-1) + \sqrt{(X-1)^2 + \frac{8klX}{cT}}\right] \tag{3}$$

d= control delay (s/veh);

d_1= uniform delay (s/veh);

d_2= incremental delay (s/veh);

d_3= initial queue delay (s/veh);

PF= progression adjustment factor;

X= volume to capacity (v/c) ratio for the lane group (also termed degree of saturation);

C= cycle length (s);

c= capacity of lane group(veh/h);

g= effective green time for lane group (s);

T= duration of analysis period (h);

k= incremental delay adjustment for the actuated control; and

l= incremental delay adjustment for the filtering or metering by upstream signals

Control delay is the best MOE describing traffic operation at signalized intersection and can also be a factor in affecting intersection safety.

V/C Ratio

One of the most important indications implying the traffic state at an intersection is the V/C ratio (1,4). V/C ratio is defined as the ratio of flow demand to the capacity of a traffic facility. This measure can be employed to determine if an existing facility has a sufficient capacity to serve the traffic demand. A desired V/C ratio should be maintained to ensure that a facility provides an adequate level of operation while serving all vehicles. It is, of course, desirable that all facilities be designed to provide a sufficient capacity to manage demands (i.e. that the V/C ratio be maintained at a value less than 1.00). The comparison of demand flow to the capacity is a critical parameter in capacity and level of service analysis of in intersection. Its effect on safety will be evaluated in this study.

Intersection Safety

The intersection is a complex location in a traffic system where vehicles have high potential to be exposed to crashes due to the right of way issue. Safety study on intersection crashes therefore attracts a lot of attention. There have been quite a few studies on intersection safety so far and many studies quantified the safety effect from factors through prediction modeling processes. From many previous studies, Volume has been mainly used as a traditional variable for explaining the nature of accidents even if it is not evident that volume alone can represent the nature of traffic accidents (2). Some researchers recognized that a traffic crash is caused by not only volume, but other factors that traffic engineers usually use to represent the traffic situation (5, 6, 7). Those traffic factors include: control delay, level of service, and v/c ratio, even though those parameter themselves are operational ones and do not convey safety measurement explicitly. The inclusion of those traffic factors is in an attempt to improve the accuracy of the model. However, few studies have been involved in left-turn crash model development using these parameters due to the lack of data related to left-turns.

Analysis on Vehicular Collision at Signalized Intersection

Many studies have been conducted to identify the contributory traffic factors for increasing the vehicular crashes at signalized intersections. They also give a great deal of effort on estimating the safety performance with regard to various traffic elements at signalized intersections. Noyce et al. (2000) (8) separated left-turn crashes in terms of four different manner of collisions as follows: *Opposing Left-Turn Conflict; Left-turn, Same-Direction Conflict; Lane-Change Conflict; Opposing Right-Turn-on-Red Conflict*. They also investigated the relationship between crash type and signal display. Of four types of crashes, the *Opposing Left-Turn Collision* accounted for around 60% of entire left-turn crashes over the targeted area. More *Opposing Left-Turn Conflicts* were observed with the leading phase than lagging phase from the collected data. This paper interpreted that driver confusion results in more crashes with leading phase due to simultaneous illumination of the green arrow and red ball or green ball. *Left-Turn, Same Direction Conflict* seems to be caused by drivers over-attentive in the gap acceptance during the permitted left-turn phase. No evidence was identified for the relationship between *Left-Turn, Same Direction Conflict* and the signal display. The other two types were very low in percentage of all conflicts and there seems to be no evidence to suggest that the rest of them are really related to signal displays. Based on the results of this study, the authors suggested that the green ball permitted indication may cause the driver's misunderstanding or confusion, and signal complexity or driver workload with simultaneous illumination of green arrow and red ball may contribute to left-turn crashes as well.

Similarly, Hauer (1988) (9) analyzed four legged intersection accidents using fifteen conclusion patterns. These patterns were determined in terms of the maneuver of two vehicles before collision. From the result of the analysis, *Opposing Left-Turn Conflict* was the highest of 15 different types. To understand how traffic flow affects the crash rate, two contributory volumes involved in *Opposing Left-Turn Conflict* were investigated along with the number of left-turn crashes. From the analysis, the author suggested that the number of accidents is proportion to the through flow. During the exploration of the given data set, the authors reached several conclusions. First, vehicular accidents are related to flows to which the colliding vehicles belong. Second, accident categorization in terms of initial impacts is not very helpful to identify the cause (accident maneuver) and effect (accidents). Third, *Opposing Left-Turn Conflict* is more related to through flows rather than left-turn approaching flows. It implies that the sum of entering flows cannot be appropriate in using for calculation of intersection accident rates when comparing the safety of two intersections.

Retting et al. (2002) (10) investigates the impact of signal timing change on changes in the crash risk. In this study, it basically showed the changes in safety performance at signalized intersections as yellow and all-red time increase. As a matter of fact, sometimes the short yellow signals do not allow drivers to get out of the intersection before the red indication appears. The authors identified that there could be potential safety benefits by increasing yellow and all-red intervals through the before and after study although some of the crash types are not very statistically significant for signal timing.

Morocoima-Black et al.(2003) (11) investigated the intersection crash trend from 1996 to 2000 to determine whether improvement of safety performance has been accomplished. This report identifies the intersections where crash experience and rate are higher than normal. The report determines criteria for selecting intersections with higher risk. Average number of crashes and crash rate are the first step to screening selected crash intersections. The authors explain the weakness of each method since crash experience does not take the approach volume into account and crash rate overestimates safety risk of the intersection having low approach volume. Thus, this report recommended that a combination of an average crash rate, average crash number, and average daily volume should be employed to achieve the most appropriate criteria.

Previous Study on Developing Prediction Models

Safety performance functions are indispensable in the safety study field because it allows recognition whether certain factors have impact on crashes for given traffic situations through such factors (e.g. traffic operation and geometry design) as an independent variable. A safety model is rare for a single urban intersection considering that the comprehensive data set is required to develop such an application model, except for models featuring volumes at signalized intersections. Lyon et al. (12) investigated collisions at signalized intersections to develop Safety Performance Function (SPF) with a data set from North America. They classified intersections into several categories depending on approach road characteristics which are road classes defined by the City. As intersections were divided, nine groups were obtained for four-legged intersections. It is very crucial to select variables that are used for a predictive model. Through a preliminary data analysis, variables were chosen that might have significant impact on crash. Beside Annual Average Daily Traffic (AADT), the authors suggested several key variables:

- single lane & multilane approach
- with and without left-turn lanes
- with and without right-turn lanes
- high and low pedestrian activity

These key variables suggested were considered as significant factors impacting on the two crash types: property damage only, and fatal plus non fatal injury combined. Best variable combinations for the predictive safety models were described through the modeling process. Among those key variables, four of them are considered as significant factors to predict crash according to different categories. The data set was used to describe how those factors affect crash type, severity, and injury level at intersections. As for left-turn crashes, the number of left-turn approach lanes on a high volume road is a significant factor for a Property Damage Only collision on only four-legged intersections. This variable has low significance for Fatal + non-Fatal Injury collisions. three-legged intersections have different combinations of variables. Safety Performance Function (SPF) coefficients were estimated by statistical software called Genstat assuming that error is described by a negative binomial distribution. Dispersion parameter related to mean and variance uses the same value along with the appropriate model. The coefficient value can be interpreted as how a variable impacts collision. For example, a three-legged intersection has -0.189 as a coefficient for a number of left-turn lanes on a low volume road in the *F+I model*. Coefficients with a negative value mean, as there are more LT lanes there is a less chance to have a collision. This study demonstrates the application of SPF using data of LT collisions at the signalized intersection in urban areas. This study handles collisions with at least one vehicle involvement. There are two types of left-turn priority treatments that were installed: flashing advanced green (FAG) and left-turn green arrow (LTGA). For 3 years, the targeted intersections recorded only left-turn and left-turn side impact collisions. The evaluation conducted was based on F+I due to erratic nature of PDO. To complete this study, the EB (Empirical Bayesian) method was used. Results from the before and after study presented that these treatments had similar impact on left-turn collisions. Based on the results, the intersections with LTGA showed a larger effect than FAG.

Bauer and Harwood (2000) (13) modeled safety performance for five-types of intersections through a statistical method. The prediction model involved geometry design as factors impacting on numbers of accidents. Design elements employed for the safety model were functional class, traffic flow, channelization, traffic control type, median, access control, terrain type, number of lanes, lane width, shoulder width, and lighting. All types of vehicle accidents were investigated. Intersections were divided into five types in terms of operation, geometry, and location characteristics. Two general methods: lognormal and log-linear regression were used in describing the accident data collected. Poisson distribution allows better describing of the data than lognormal when discrete events occur at intersections. Although log-linear equation is similar to lognormal, the latter equation was developed based on the Poisson assumption and mean μ. Poisson distribution has limitations that the mean and variance should be similar. Due to that limitation, the negative binomial distribution also applies to the accident data. (14). A negative binomial has two parameters: the dispersion parameter k and mean μ to describe the data. The Logarithm form is used for the Average Daily Traffic (ADT) count for each direction.

The accident prediction models were developed for each of the five intersections. For log-linear, the negative binomial is preferred to the Poisson distribution due to over-dispersion. ADT of both major and crossed roads accounted for almost all the variability in the accident data described by the model. Geometric designs were found not to account for significant portions of the variability.

For all types of intersections investigated, the geometric design features were statistically significant in negative binomial regression models which included number of lanes on major roads, presence of major-road left-turn prohibition, type of access control on a major road, width of a major-road outside the shoulder, presence of a major-road and crossroad right-turn channelization, design speed of a major road, presence of a median, presence of a major-road left-turn channelization, and average lane width on a major road. Types of terrain, functional class of major road, and presence of lighting at the intersection were also statistically significant. A ratio of Pearson chi-square to degree of freedom and deviance to degree of freedom were used as criteria of goodness-of-fit tests. From the results, the negative binomial turned out to be appropriate to describe the number of accidents, and number of fatal and injury accidents. The equation below is involved in the urban four-leg signalized intersection. The prediction model of that has the figures like the following expression(12):

$$C = 0.001066 Q_{major}^{0.574}\, Q_{\min or}^{0.215}\, e^{B_1 + B_2} \tag{4}$$

$$B_1 = -0.051 I_{pt} + 0.400 I_{fa} - 0.240 I_{mp} - 0.290 I_{ac} \tag{5}$$

$$B_2 = -0.155 I_{\min} - 0.163 I_{maj3} - 0.151 I_{maj4} + 0.005 V_d \tag{6}$$

C = frequency of severe intersection-related crashed, crashes/yr

I_{pt}= control type indicator variable (1 if intersection is pre-timed; 0 if semi-actuated)

I_{fa}= control type indicator variable (1 if intersection is fully actuated; 0 if semi-actuated);

I_{mp}= signal phasing indicator variable (1 if more than 2 phases; 0 otherwise)

I_{min}= minor-street through lanes (1 if 3 or fewer lanes; 0 otherwise);

I_{ac}= major-street access control indicator variable (1 if no access control; 0 if partial control);

I_{maj3}= major-street through lanes (1 if 3 or fewer lanes; 0 otherwise);

I_{maj4}= major-street through lanes (1 if 4 or 5 fewer lanes; 0 otherwise); and

V_a= major-street design speed (mph)

As the equation shown above, control type, number of phasing, and number of lanes for the minor street can affect the crash frequency. Pre-timed controllers have fewer number of accidents than semi-actuated, and fully actuated controllers contribute to higher number of crashes than semi-actuated controllers. Major roads with no access control have fewer crashes compared to ones with access control. The model shown above does not provide any reason or justification; it is just an intuitive trend. Lastly, an expectation for the model does not have high goodness-of-fit. Thus, this model could be used as a reference for future work. However, it cannot be applied to the real field.

The National Cooperative Highway Research Program conducted a safety study to estimate the impact of signal on safety performance at an intersection (15). MUTCD provides safety warrants for installing signals at intersections; however, it does not require intersections to install signals. The objective of this NCHRP report is to determine whether signal impacts are statistically significant for the safety-performance at an intersection. An accident count comparison on one with signal and one without signal by using a control group was established. This report conducted calibration of the crash prediction models to show how signal impact on intersection safety. The process of developing the safety model involves exploring variables to which variable have significance in increasing or decreasing in accidents. A negative binomial distribution was applied to describe how accidents distributed at a targeted intersection. Dispersion parameter, k, was employed to find a model that has a better fit for the given data and the relationship between the mean and variance can be explained by using ‘k’. The higher the k value is, better the result will be in terms of lower variance. Only volumes were used as variables in the safety model. This report presented two reasons for using only volume as the variable: one is other variables such as sight distance and approach speed are not sufficient to explain accident variations for different intersections. In addition, volume can explain much variation in an accident as the previous research. Genstat was employed to calibrate coefficients for each variable and the intercept. The safety prediction model was developed for each turning movement at the targeted intersection.

M. Abdel-Aty et al. (2005) (16) were trying to identify how a geometry design and crash-specific aspect have impact on the injury level of crash occurring at signalized intersections. The level of severity for accidents was classified into five groups. The ordered probit model is popular for the description of the dependant variable’s ordinal nature. They used the ordered probit modeling technique to identify variables impacting significantly on the level of severity for accidents occurring at signalized intersections. The data consist of crash severity that was used to construct ordered probit models. A total of 21,204 crashes were used for this study. This research has two objectives comprised of roles exposing factors which effect crash severity as well as to verify if the probit models could produce different output depending on the types of accidents. Results from the ordered probit model associated with the crash type presented that the complete model gives better results than a restricted model based on classification. The seven crash types were recreated by dummies using the complete model. The result indicates a coefficient that shows how certain types of crashes contributed to crash severity at a signalized intersection. Of those crash types in the study, Left-turn crashes led to a high severity except for the accidents involving bicycles and pedestrians. The combined model identified left-turn, head on, and angle as the most significant factors on crash severity. The model associated with intersection characteristics was presented as well. Due to the classification accuracy, the complete model yielded much reliable results compared to the restricted model. The models presented that the number of lane and speed limit on the minor road has a role of decreasing the crash severity whereas the number of left-turning lanes and traffic volume on the major road increased the crash severity. The authors also applied a Hierarchical Tree-Based Regression (HTBR) to determine significant factors for each crash severity level. It has several advantages over the ordered probit models that demonstrate factors for the overall severity level. One of the most important benefits of the HTBR is that there is no assumption of the population's functional form. It also has the robustness against multicollineaity among factors. The ordered probit models and HTBR do not even have similarity in the results except for that the speed limit on the minor road is significant for lower injury severity levels in both methods. The authors concluded that it is crucial that each severity level should have different models correspondingly instead of using an aggregate model when expecting a number of crashes of different severity levels.

The left-turn lane effect on safety performance has been demonstrated differently in the past study. Endogenous problems are studied on installation of the left-turn lane from time to time. Kim et al. (2006) (17) examined endogeneity problems by using a crash model at a signalized intersection. The data set in this research has been used beforehand to determine if left-turn lanes have different effects on intersection safety when endogeity problems were considered. Two models were developed to demonstrate if one controlling endogeneity problems can have a different result from one without control. To prove this, a negative binomial model was developed for an angle crash and a logit model for a left-turn lane. There are five and four significant variables for forecasting future safety performance at the intersection for binomial and logit models respectively. Each model is utilized to estimate how independent variables affect each corresponding dependant variable. Afterwards, Limited information maximum likelihood (LIML) was used to ensure if endogeneity impact is significant for a safety model. Based on the results, the negative binomial models showed that the presence of a left-turn lane is significant to a number of angle crash without control. This is one contrary to engineers' intuitive thought; whereas the presence of a left-turn lane has negative impact on a number of angle crashes by models with controlling endogeneity by LIML. In case of the logit model that has angle crash as an independent and a left-turn lane as a dependent variable, it presented a larger number of crashes on left-turn lanes without endogeneity control. However, the angle crash is an insignificant variable when a model can control endogeneity by an LIML approach. Lastly, a test was applied for endogeneity to verify if the model can be consistent without it. As a result from this test, models can not have consistency without an endogeneity control according to the p-value.

There have been several studies associated with highway access. Among many researchers working on this topic, Garrett D. Burchett et al. (2005) (18) identified the factors associated with geometry design and concluded that the land use impacts the safety on the highway sector. First of all, five severity indexes were defined to categorize all the accidents. Then, they investigated 100 intersections with best safety performance and 100 with worst safety performance out of the entire intersections by a statistical comparative analysis. Descriptions of the interaction between safety performance and geometry design, and safety and land use was done by using a database. Four types of geometric features categorized the examined intersections. The three different land use types were applied to describe the intersection characteristics. Each different characteristic defined by the author had significant interaction with accident severity at the intersections. The negative binomial model was used to construct the safety performance function (SPF) due to the reason the crash data in the dataset has over-dispersion. This model describes how those independent variables impact accident severity and frequency. The final model was determined through a Rho-square value and parameter estimates out of every potential model. Rho-square indicates how the model developed is appropriate to describe the dataset. Statistical significance in parameter is to show how a certain independent variable is reliable on forecasting future performance yielded by the model. From this process, the final model for SPF presented that two out of four geometries designs: intersection skew, vertical curve, affect the crash severity. Commercial area was identified as positive impact on crash severity.

Further research was conducted on 30 intersections with the poorest safety performance besides descriptive and statistical analysis. Different time periods (peak hour and off-peak hour) were used. Two different crash data sets in the corresponding time periods were compared to identify how volume impacts crash. When extremely high flow exists, intersections experience the highest crash count. The distance between the major road and the minor road was also examined to identify the relationship between the accident and intersection geometric features of interest. The far lane from the minor approach road has a higher crash experience. The authors insisted that geometric features attributed to the crash count. Although the sample size is low, the data showed relationships between the examined factors.

LOS (Level of Service) has been focused to determine how traffic safety would be influenced at a certain LOS. Bhagwant Persaud and Thu Nguyen (19) reviewed the past research of the effect on a congested traffic flow on a roadway. The statement was made by the reviewed research, where as the roadways become more congested the level of safety also worsens. It is not a surprise judging as how accidents are more prunes to occur as more interactions between vehicles exist on the traffic facility (i.e. ramp, highway, intersection, and collector road). The probability of crash may increase as implications of roadways have breakdown or bottleneck situations. The first step to develop a safety model was employed by a basic safety modeling form. Many researchers used the following model form:

$$P = \alpha S^{\beta} \tag{7}$$

where S is the traffic volume, and both α and β are calculated parameters. There could be several other variables besides volume. However, an attempt was made to only add a level of service indicators as a variable. Although there would be correlation between volume and level of service, the level of service indicator was involved in a safety model as a significant variable for that correlation is not as critical as in predicting a model as it has an impact to the future forecast safety performance. The following is the example of the model form:

$$P = \alpha S^{\beta_1} Y^{\beta_2} \tag{8}$$

Y can be a level of service indicator such as delay, capacity and v/c ratio. These two equations are non-linear and P (accident count estimation) follows a negative binomial distribution. In order to facilitate calibration of the parameters (α and β), the GLIM software package was used with a negative binomial error structures. This study determined if the models developed is reliable in terms of dipersion parameter, *k*. It is similar to R-square value which is a statistical measure of model fit. However, the R-square value may not be appropriate in a generalized linear modeling of variables with a non-normal error structure.

$$Var(P)=P^2/k \tag{9}$$

As the equation shows, variance is larger with a smaller *k* value and vice versa.

Only peak period volume data were used instead of a total volume such as AADT or ADT. All data were treated individually by impact type, accident pattern, and corresponding traffic flow except for the aggregate data such as total volume and number of accidents. The average value of each two-hour rush period volume and appropriate peak hour factor (PHF) were applied to calculate average stopped delays, capacities, and level of service using HCS. Two models were developed according to the number of approaches for intersections: four-legged intersections employed delay time as a level of service indicator and three legged intersections used dummy variables as a service indicator. According to the *k* (dispersion parameter) value, two models are most fit to the data as shown in Table 1:

TABLE 1 Regression model for intersection collision

Four-legged intersection	Three-legged intersection
Model form	
$P = \alpha S_1^{\beta_1} S_2^{\beta_2} Y_d^{\beta_3}$	$P = \alpha S_1^{\beta_{maj}} S_2^{\beta_{\min}}$
S_1 is the flow on the major approaches S_2 is for the minor approaches Y_d is the average stopped delay(sec)	β_{maj} and $\beta_{\min}$ are exponents for the mojor and minor flows, both α and $\beta_{\min}$ depend on the LOS category..

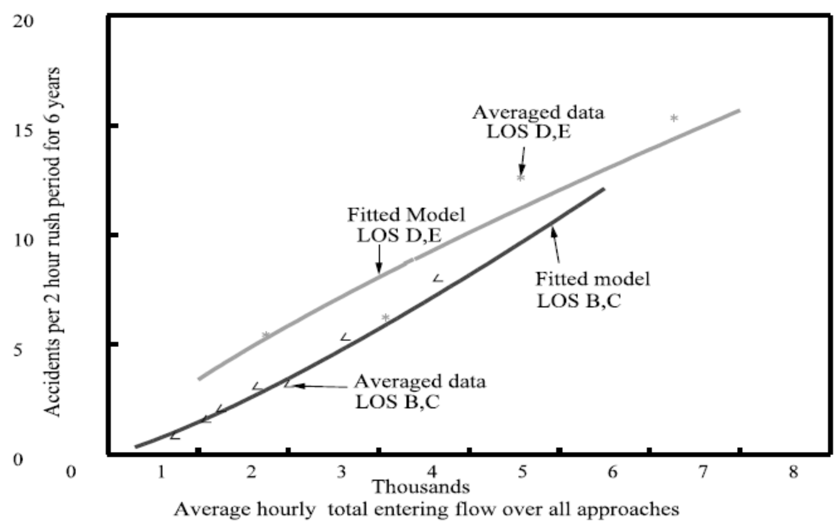

FIGURE 3 Regression lines and averaged data for model B2

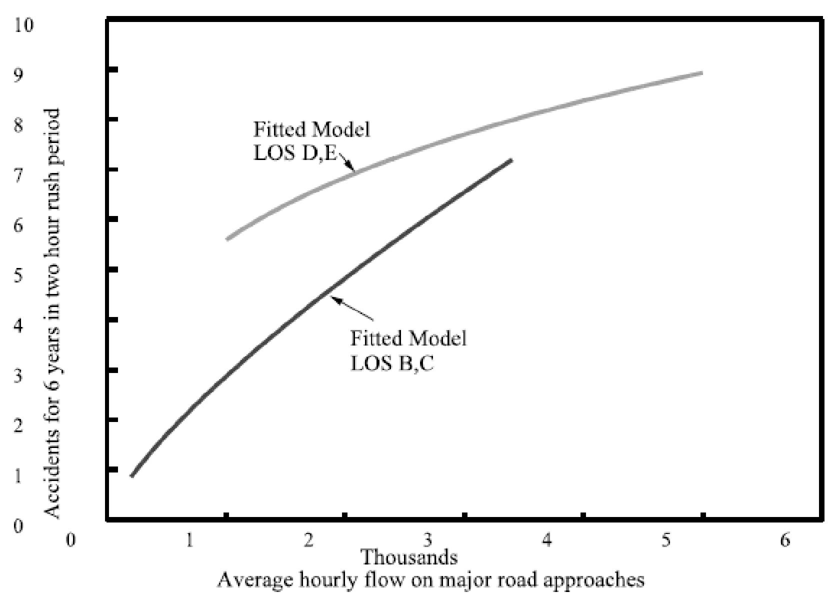

FIGURE 4 Regression lines for model B4

Figures 3 and 4 show the plots of predictions from each model that covers the appropriate range for the flow. It also describes the difference of accident counts between the two LOS categories. This discrepancy appears both on an average data and fitted model data. The authors concluded that the LOS indicator can be utilized for forecasting a future safety performance at the signalized intersection. There also can be an implication that it might be useful in the planning step for decreasing accident counts.

Ching-Yao Chan (2006) (20) examined left-turn accidents at an urban signalized intersection. Left-turns are critical movements which may have drivers exposed to a higher potential of crash at signalized intersections. Around 40% of accidents were related to left-turns according to the data. Permitted left-turns always have a conflict point with opposing volume. In this study, a left-turn vehicle was called a subject vehicle (SV) and the opposing vehicle is called a principle over the other vehicle (POV). The author investigated left-turn potential conflicts with regard to when a left-turn has permission and identifies where to put a conflict warning device informing the vehicle that there could be a potential conflict before the SV takes the left-turn. The time difference between the SV and the POV is most crucial in a potential crash. The place that is likely to have a crash between an opposing through vehicle and a left-turning vehicle is referred as a conflict zone. A left-turn accident usually occurs in a conflict zone caused by a time difference between the POV and SV. In other words, if the POV passes the conflict zone after the SV crosses that area, it could bring a higher and more severe crash. On the contrary, if the SV is holding and waiting until the POV passes the conflict zone, the time difference between the two vehicles could be very small. A former case is called the *trail buffer* and there could be a much higher potential conflict under this situation. There could be the case that the POV slows down as it approaches the intersection and takes a right turn. If the SV expects to cross the conflict zone before the arrival of the POV, and they are heading for the same destination, then it would become a potential conflict. With this insight, three different intersections were investigated to find when the SV interacts with the POV with a higher potential conflict in those targeted areas. Each site has different characteristics such as heavy vehicle, signal operation, speed limit and pedestrian portion except for the left-turn type. All of them have permitted left-turn so that interaction between the SV and the POV could be evaluated. The following four major attributes were determined:

1) *Driver Factor- poor perception, judgment, or physical capacity*
2) *Pedestrian- pedestrian presence leading to a threat to the SV*
3) *Signal Transition- SV maneuver taking place during signal phase change*
4) *Informed Decision- driver decision to proceed under non-threatening condition*

From the field data observation, each intersection has unique characteristics according to their traffic situations. Driving behavior had the largest significance of the entire case based on overall analysis while pedestrian and signal transition had a low impact. The author listed some revealing characteristics of potential conflicts based on the data analysis.

1) Pedestrians cause to interrupt drivers and may lead vehicles to risky situations.
2) Higher speed and dense-traffic flow result in more frequent conflicts.
3) Drivers attempt to initiate turns in the late green or amber. This behavior can lead drivers to potential conflict. A CICAS(Cooperative Intersection Collision Warning System)/IDS(Intersection Decision Support) system should be installed to the place to avoid unnecessary alerts.
4) There could be potential conflict when the POV prepares to take a right-turn, and the SV thinks that the POV would slow down to stop. To prevent this situation, a CICAS/IDS needs to differentiate this case to minimize annoyance alarm recognized by the SV drivers.
5) The place which has high percentage of left-turn exposed to a conflict potential should be classified to the situations where warning needs to be warranted to drivers.

The author concluded that the most significant situation for potential conflict can be determined on each intersection in terms of 4 factors. Based on this information, the place needs implementation of CICAS/IDS for providing drivers with assistance to increase the ability to recognize and avoid potential conflict situations before exposure to the situation.

Lord et al. (2005) (2) investigated relationships between the accident and traffic flow that has been conducted in a traditional way. The traditional way to build a safety model has been mainly focused on the traffic volume in order to forecast future safety performance in entire road way segments even if the volume could be aggregated or disaggregated. However, this is not very clear to identify the traffic flow as a main factor affecting the crash frequency. Besides traffic volume, other traffic characteristics such as V/C ratio and density should affect the accident count over the targeted road segment. With this concern, the authors examined how traffic characteristics have impacts on roadway safety with two sections' data: urban and rural that has already been used for a previous study. The crashes were divided into three categorizes in terms of number of vehicles involved in a crash. The authors investigated field data before developing statistical models. In the rural sector, the increase of density and V/C showed the decrease of single vehicle crashes. On the contrary, multi-vehicles have an opposite trend. Accident severity increased with V/C but not very related to density. For urban areas, the increase of density increased the total number of crashes. For predicting future safety performance over selected areas, safety models were developed using a negative binomial structure. To evaluate the developed model, traditional methods were not used due to several reasons that the models are not nested and a dispersion parameter is not fixed. Furthermore, the mean is very low in both datasets. The deviance estimation and cumulative residuals were applied to evaluate the model. The resulted models involving other characteristics as covariates yield a better fit than the model with only traffic flow.

The authors illuminate several points to be discussed in this experiment. First of all, the predicted model solely incorporating traffic flow may not have a better fit than the model with other traffic characteristics. For instance, the field data show that the numbers of accidents decrease as the number of vehicles increase. This indicates the traffic flow may not be adequately representing the nature of accidents. The vehicle density and V/C ratio showed the trend what one would expect for in the real world.

Secondly, one predicted model cannot be used for all crashes on a highway segment. The model should be developed separately as a single-vehicle and multi-vehicle crash for obtaining more adequate results. The functional form is another important issue. Lord et al. suggested that a log should be used for the volume to be more rationale even if the volume has a linear relationship with accidents as well as a log form. A few models have a negative coefficient for volume; whereas traffic parameters such as a V/C ratio and flow density have negative coefficients as well. In other words, traffic flow characteristics might have different roles under different traffic situations. The authors stated that a better fit does not always explain the accident characteristics over targeted areas. In sum, all the factors involved in traffic characteristics should be considered in developing the safety model for a better illustration of accident characteristics. With the better understanding of traffic accident's traits, a better design for highway segments could be established in the planning step.

B. Persuad and T. Nguyen (21) identified risky intersections and estimated the safety performance of a signalized intersection based on an Empirical Bayesian process. Two different levels of models have been developed for this objective due to the amount of available data. The models for safety performance were developed for three-leg and four-leg intersections by accident severity (injury and POD) at Level 1. The models considered all three major-crash types for calibrating equations. Different models were developed separately for different time periods. At Level 2, the prediction models were estimated for four-legged intersections by different crash patterns: 12 multi-vehicles and 3 single vehicles. Those patterns described an accident involving vehicle movement before a collision. The Empirical Bayesian method was employed to expect the future number of accidents at the intersections. This study solely used traffic flow as an independent variable because the data quality of other variables is not sufficient enough to satisfy the theoretical pattern. They employed two different model forms for each levels.

For Level 1 and for Single Vehicle 2: $P = \alpha S^{\beta}$

For Level 2 Multivehicle: $P = \alpha S_1^{\beta_1} S_2^{\beta_2}$

GLIM was used to calibrate α and β for safety performance at the intersections. Log-link function was utilized for computing α and β. The model form is transferred below:

$$\ln(P) = \ln(\alpha) + \beta \ln(S) \tag{10}$$

GLIM provides a standard error of estimates of ln(α) and β. Since ln(α) has a very low standard error compared to β, only the standard error of β is reported. The model for Level 1 used a sum of entering volume of the intersection approach and accident types combined and separately due to sufficient accident data. Models in Level 2 for a single vehicle almost have the same approach as models in Level 1. Level 2 models for multi-vehicle accidents, only 4-legged intersections were applied to build a prediction model and 25 accident patterns were defined by a vehicle movement before a collision. As defined before, a non-linear prediction model form was $P = \alpha S_1^{\beta_1} S_2^{\beta_2}$. A log-link function should be used for calibrating α, β_1, and β_2. S_1 and S_2 are referred as AADT of primary and intersecting road respectively at the targeted location. Owing to the limitations of available accident data, all types of severity data were combined for modeling which is contrary to level 1.

Iowa State investigated the impact of a left-turn type on older and younger drivers (22). The left-turn type is usually decided in terms of reflection on the operation. There are quite a few studies that have shown the impact of a left-turn type on intersection safety. The safety model was developed to determine the interaction between the crash frequency and independent variables. (i.e. age, left-turn type, total opposing ADT) As for age, the drivers were spitted into three different groups. 7 variables were used as variables based on the Poisson process. The categorical variables were used to build the prediction model written as:

$$CrashRate_{left} = e^{(-0.8753 + 0.18 prot/per - 1.6311 prot - 0.4454 Age0 - 1.0823 AgeI - 0.6158 Opp\_AADT)} \tag{11}$$

Where,

$CrashCount_{left}$ =	Left-turn crash count (crashes/MEV)
Prot/Per=	Dummy variable for protected/permitted phasing (0 if phasing is protected or permitted, 1 if phasing is protected/permitted)
Prot=	Dummy variable for protected/permitted phasing (1 if phasing is protected or permitted, 1 if phasing is protected/permitted)
Age0=	Dummy variable for age group (0 for 25-64 and 65+ and 1 for 14-24)
Age1=	Dummy variable for age group (0 for 14-24 and 65+ and 1 for 25-64)
Opp_AADT=	AADT for opposing approach (1 for AADT from 5000 to 9999; 2 for AADT from 10000 to 14999; and 3 for AADT from 15000 to 19999)

From the modeling process, the expected crash count in terms of a protected left-turn is lower than a permitted left-turn. Likewise, a permitted plus protected left-turn is higher than a permitted left-turn. According to the estimated coefficient, this model also illustrates that the oldest group has the highest expected crash. As oppose to what one would expect, the opposing ADT contributes to reducing the crash count.

Summary of Previous Safety Studies on Left-Turn

In quite a few previous studies, intersections have been treated as a critical space for safety issue in traffic systems. Many have developed crash models using statistical methods and identified statistically significant factors during the modeling process to identify traffic elements affecting the change in intersection safety. Among all such identified factors, traffic volume features prominently, although it is not evident that it can fully explain the nature and frequency of traffic crashes. Some researchers have recognized that crashes are caused not only by volume, but can also be attributed to other traffic and roadway characteristics, which includes control delay, level of service and volume over capacity ratio(2, 21). Inclusion of these traffic factors is based on the assumption and intuition that there should be correlation between those factors and accident counts. Although many researchers accomplished the safety prediction model using covariates associated with a left-turn, these models are not concerned with left-turn crashes. Most studies regarding safety at all approaching movements for intersections are recorded no matter what the cause, but this study mainly focuses on left-turn crashes. Very few studies have focused left-turn crashes(19, 21). For this reason, prediction models for left-turn crashes were motivated to be developed.

CHAPTER III

DATA COLLECTION AND DESCRIPTION

Data were categorized into two parts: crash and intersection data. Crash data were collected from 2002 to 2004 by the College Station Police department. The crash data basically includes location, manner of collision, and time etc. Intersection data have geometry design, approaching volume, post speed limit and signal timing data. These data were utilized to compute V/C ratio and control delay using HCS.

Crash Data and Characteristics

Accident data collected in the City of College Station are described in very good detail. Each accident record described characteristics such as manner of collision, date, time, weather condition, road condition, accident cost, location, severity, and the driver's name for the corresponding accident etc. However, many of these characteristics were missing. Since this study focuses on left-turn safety, left-turn accidents at signalized intersections were extracted out of the entire accident. Then, they were categorized by three time periods corresponding to volume. The location where accidents occurred was indicated by using primary and intersecting roads. Although accidents occurred at the same intersection, it could be treated as a different case by approach, accident time, and volume corresponding to the time period. The crash data is collected by three different time periods: morning, noon, and evening at the same location. As Figure 5 shown, most of the intersections in the targeted area have 'zero' crash experiences for three years. The mean number of crashes is very low and the variance is large according to collected data. This phenomenon is a well known characteristic of a crash data. Of 115 observations, 69 were not involved in crashes. Only 40% out of the entire dataset have crash experiences. In addition to this small number of crashes, it is assumed that all crashes occur independently. Rarity and independence of crash data are a feasible characteristic for using a negative binomial. Basically, Table 2 has the same information as in Figure 5.

TABLE 2 Number of left-turn crashes vs. observations

No. of Crashes	0	1	2	3	4	6	9
No. of Observations	69	23	10	6	5	1	1

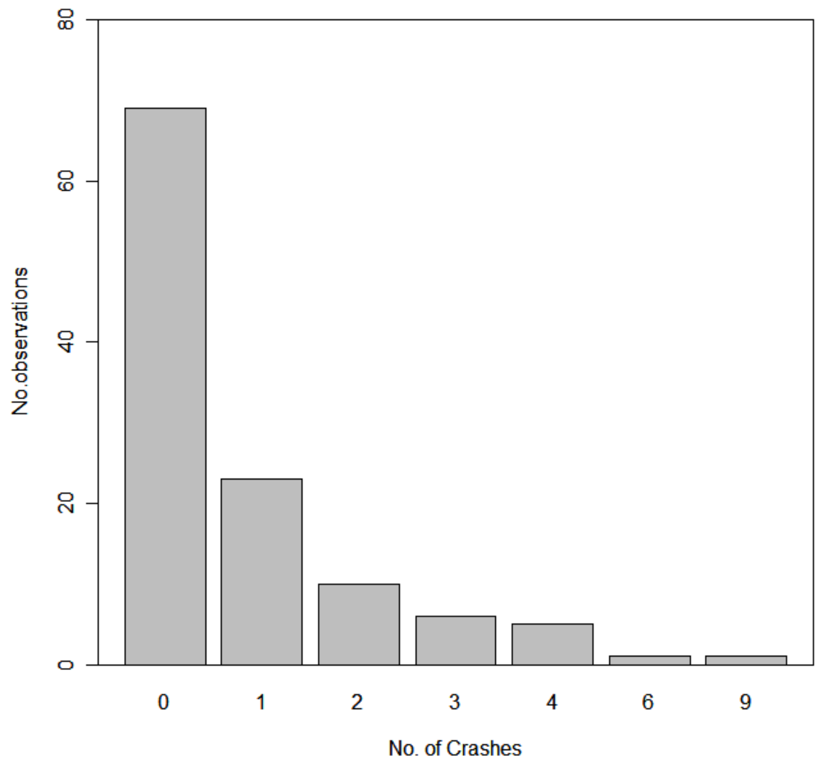

FIGURE 5 Left-turn accident distribution

Volume

Intersections have traffic volume data of AM-peak, NOON peak and PM peak period (hourly volume). Hourly volumes were collected from **7:30 to 8:30**, from **12:00 to 13:00**, and from **17:00 to 18:00** respectively. The volume data as shown in Table 3 were collected every 15 minutes. Each approach, such as southbound, northbound, etc., is divided into three movements: through, left-turn, and right turn. PHF is available for each direction. There is no vehicle type information for the given data set. Volume has been generally used as variables for a prediction model even if volume is not able to explain the nature of the accident (18).In most of the safety models, both primary and intersecting road volumes have been used as variables for forecasting safety performance at signalized intersections. It is assumed that volume can be explained by variations in crash. The accident data do not have exact location information. The crash data set provides primary street and intersecting street names. Due to this condition, there is no way to figure out the directionality of the left-turning crash vehicle. Therefore, volumes were aggregated by both directions.

TABLE 3 Example for volume data

University Drive						
	Southbound			Westbound		
Interval	Right	Through	Left	Right	Through	Left
4:30-4:45	39	137	48	39	135	65
4:45-5:00	35	138	57	33	165	77
5:00-5:15	28	174	56	46	213	88
5:15-5:30	40	151	60	42	161	96
5:30-5:45	38	162	62	34	165	87
5:45-6:00	46	192	56	35	161	75
Pk Hr Vol	152	679	234	157	700	346
PHF:	0.83	0.88	0.94	0.85	0.82	0.90
Total		1065			1203	
	North Bound			South Bound		
Interval	Right	Through	Left	Right	Through	Left
4:30-4:45	75	126	83	127	197	33
4:45-5:00	68	107	67	147	188	32
5:00-5:15	67	114	88	184	216	35
5:15-5:30	91	136	102	184	184	25
5:30-5:45	69	134	102	182	175	29
5:45-6:00	54	118	99	153	123	16
Pk Hr Vol	281	502	391	703	698	105
PHF:	0.77	0.92	0.96	0.96	0.81	0.75
Total		1174			1506	

Signal Timing Data

This timing data are from public works in the City of College Station. Each intersection's signal timing data are available for 24 hours. Also the signal timing plan is changed by specific periods. In other words, signal timing is changed by schedule. The weekday schedule will be utilized as default signal timing. Some of the intersections have fully actuated signal timing plans. Only the basic phase plan is available. Only 6 intersections are fully-actuated for signal timing. These intersections have only basic information such as minimum green, maximum green, passage time, and yellow/red change. Those 6 intersections have been excluded from the dataset. All intersections involved in this research have actuated-coordinated signal timing plan during the periods of data collection. A signal timing plan is rarely changed according to the City of College Station. A sample signal timing plans obtained from the City of College Station is shown in Table 4:

TABLE 4 Signal timing data

TX ave vs. Univ								
Phase	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Min G	8	8	8	8	8	8	8	8
Max	40	60	45	65	30	60	35	70
Y	4	4	4	4	4	4	4	4
R	2	2	2	2	2	2	2	2

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**cycle**	**offset**
AM Peak	29	34	16	41	16	47	16	41	120	94
Off Peak	25	33	17	35	19	39	20	32	110	102
PM Peak	28	35	31	36	16	47	21	46	130	115

Left-Turn Type

There are three types of left-turns: permitted left-turn, protected left-turn, and protected plus permitted, used at various signalized intersections. In order to identify the relationship between the left-turn type and crash count, several diagnostic tests were conducted in an attempt to find clues using a statistical method. Out of 115 observations, 69 are protected left-turns, 36 are protected plus permitted left-turns and the rest are permitted left-turn. Figure 6 illustrates the crash distribution of each left-turn type. Half of the permitted left-turns and protected left-turns have no accident experiences. Both Figure 6 and Table 5 show that protected plus permitted left-turn seems to be the most risky left-turn phasing type. A protected plus permitted left-turn is a type of a left-turn signal phasing designed to minimize the exclusive left-turn phase time. With a protected plus permitted left-turn, vehicles may face a yellow trap problem at signalized intersections if lead lagging lefts are used. The yellow trap condition basically makes a left-turning driver attempt to get into the intersection when it may be a safety risk. When the permitted left-turn movements in one direction is changed into lagging protected movements in one direction, a left-turning vehicle from the permitted phase from the lead left direction could be exposed to a yellow trap situation.

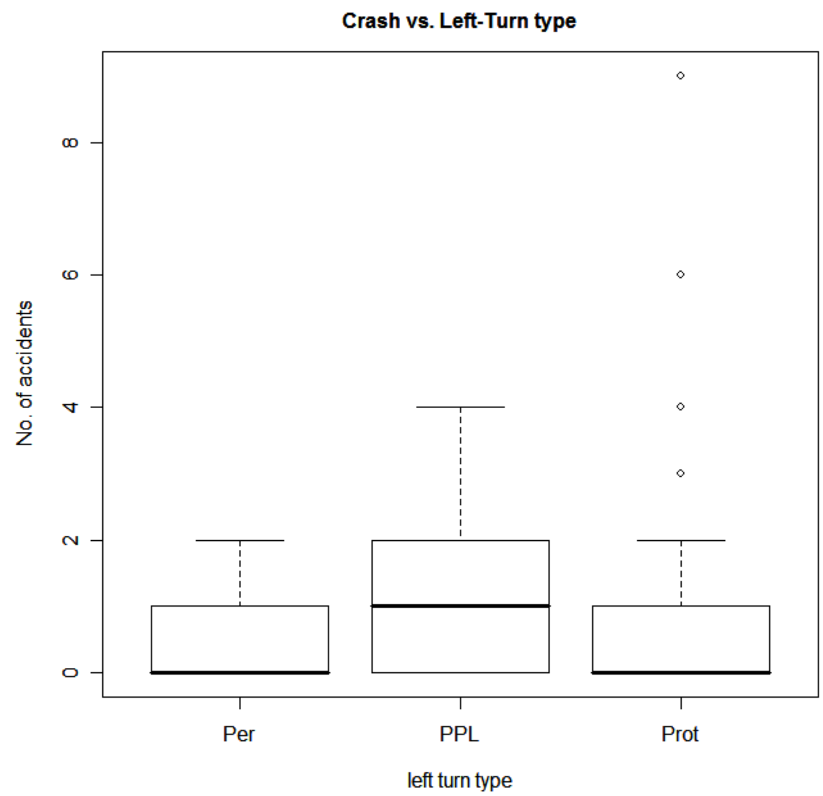

FIGURE 6 Box plot of the number of accidents for left-turn type

TABLE 5 left-turn phasing type vs. average number of crash

Type	**Permitted**	**Protected/permitted**	**Protected**
No. intersection	10	36	69
Average number of crashes per site	0.6	1.22	0.67

Left-Turn Phasing Sequence

A left-turn sequence was decided by whether or not a left-turn starts before the through movement in one cycle. Figure 8 shows that leading lefts is the most widely used phasing sequence type in the dataset. Lead-lag left-turn sequence has 29 observations out of the whole dataset. Figure 7 provides clues that the lead-lag sequence shows left-turning drivers facing more dangerous situations involving in left-turn crashes even though it is not the most widely used in this area. There is no doubt that the lead sequence has the largest number of left-turn crashes from Figure 8 and Table 6. Lead-lag has the second largest number of left-turn crashes. It seems that lead is more dangerous than any other sequences based on crash number alone. However, the total number of intersection with lead left operation should be considered in identifying relationship between left-turn phasing type and accident. As the matter of fact, the total number of cases with lead sequence operation should be considered. In table 7, lead-lag has the highest mean value for crashes compared to the other type of sequences.

TABLE 6 Number of Crashes vs. Phasing sequence type

No. Crashes	**0**	**1**	**2**	**3**	**4**	**6**	**9**
Lag	2	0	1	0	0	0	0
Lead	38	11	4	4	3	0	0
Lead/Lag	14	7	3	2	2	0	1
Split	15	5	2	0	0	1	0

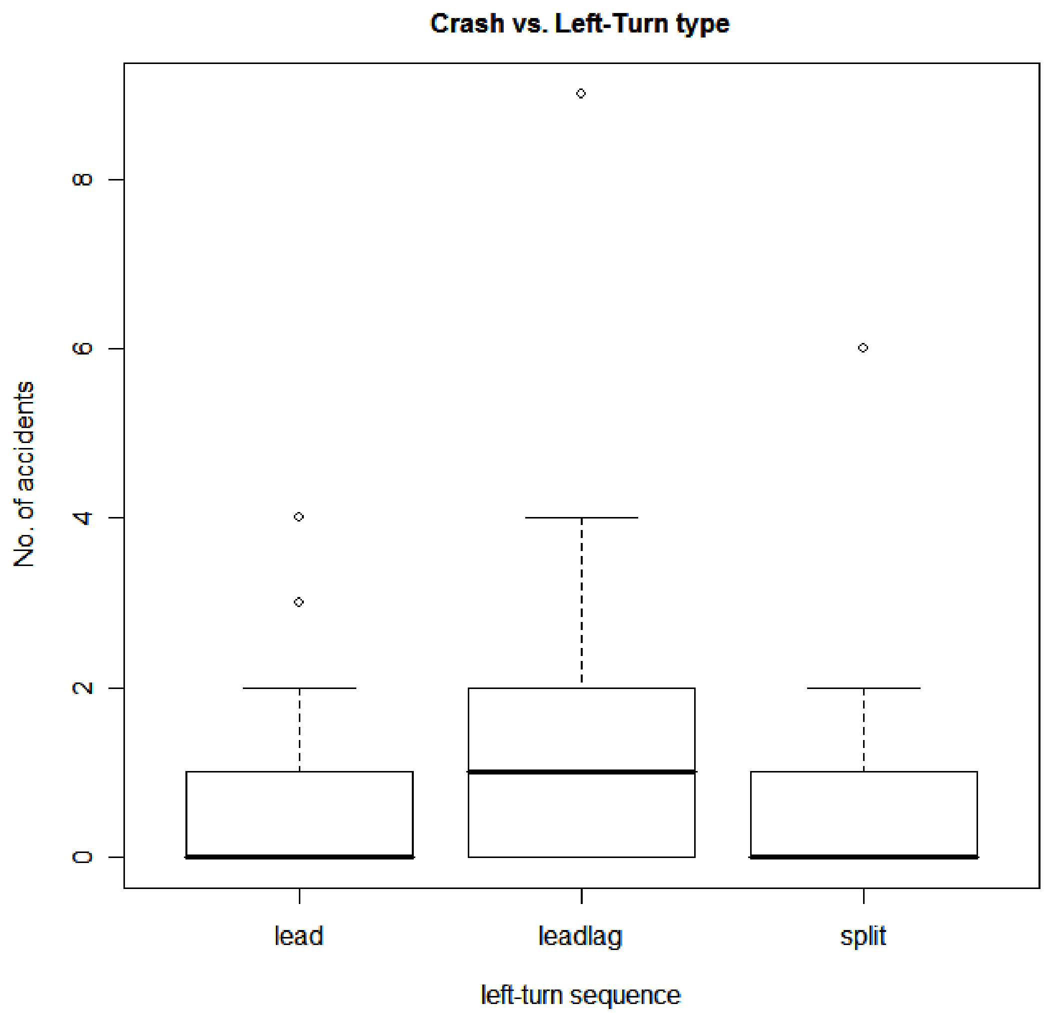

FIGURE 7 Box plot of number of crashes for each phasing sequence

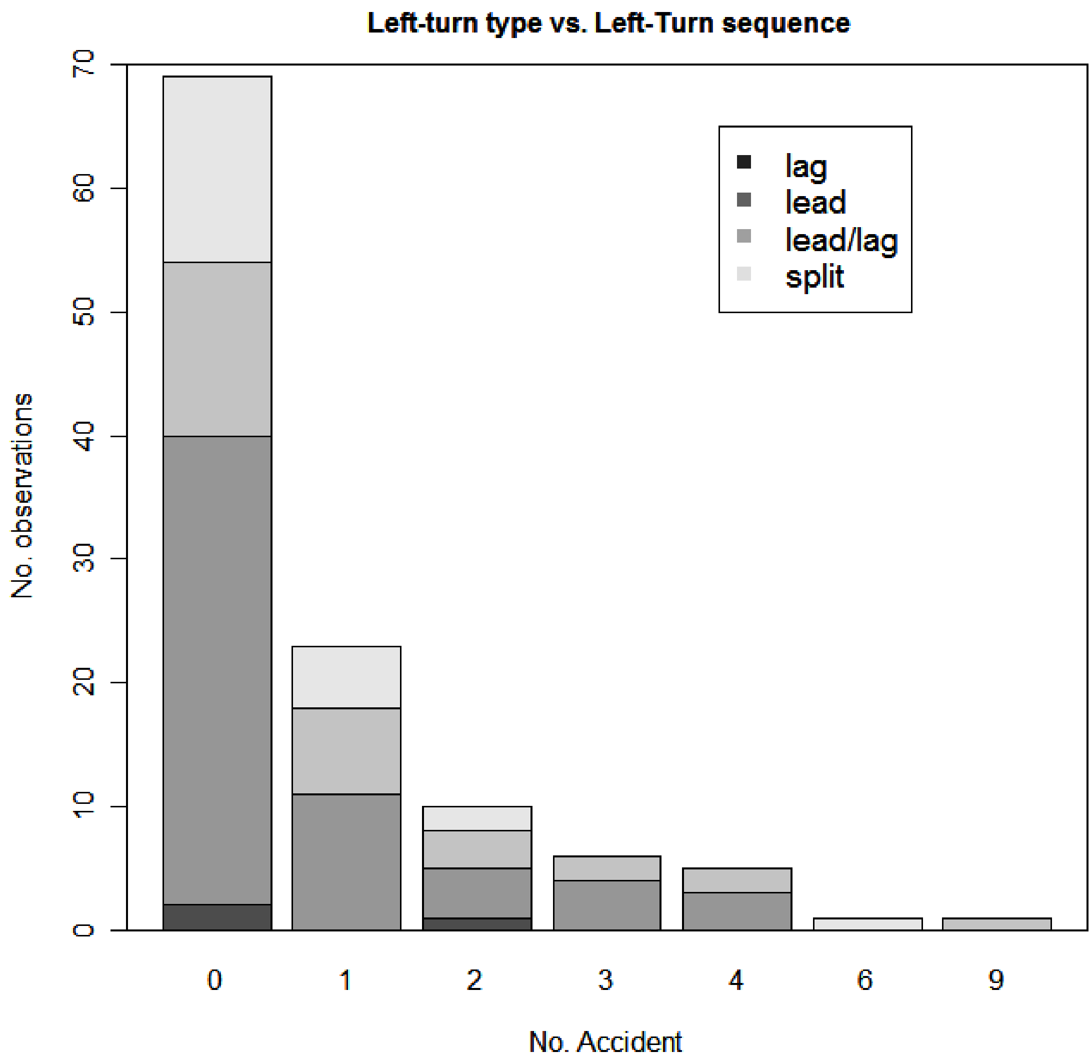

FIGURE 8 No. observations for Left-turn phasing type and phasing sequence

TABLE 7 Left-turn phasing sequence

Left-Turn sequence	lag	Lead	Lead-lag	Split
No. observations	3	60	29	23
Average number of crashes per site	0.67	0.65	0.96	0.65

Post Speed Limit

Speed is an important traffic element to be considered in any areas of traffic analysis such as operation, planning and safety. Drivers should have very short reaction times to make a decision while driving at high speed because an urban arterial may have many options for drivers to decide. As for left-turns, this condition would be very dangerous if they make a left-turn at higher speed. It would be worse if drivers are not familiar with geometry characteristics of the intersections. Furthermore, the severity of crash would become worse if the vehicles enter the intersection at a higher speed. As presented in previous data descriptions, since there is no directional information associated with left-turn crashes, average speed limit was used as a variable in this study. Figure 9 shows that largest number of left-turn crashes occurred when average speed is from 35 to 42.5mph. It may imply that most of vehicles run at those speed range.

During the data collection, several significant limitations were recognized. One of the most significant limitations is that crash data do not indicate the directional information of the crashes. For this reason, averaged values of the two directions were used in the study for traffic factors such as control delay, V/C ratio, volume, and speed. This aggregation might lead to model bias or inaccuracy for predicting and estimating left-turn crashes. In real-world, many traffic variables may have distinct directional distributions. For example, during a rush hour, volume from one direction could be quite different from the volume of the other direction.

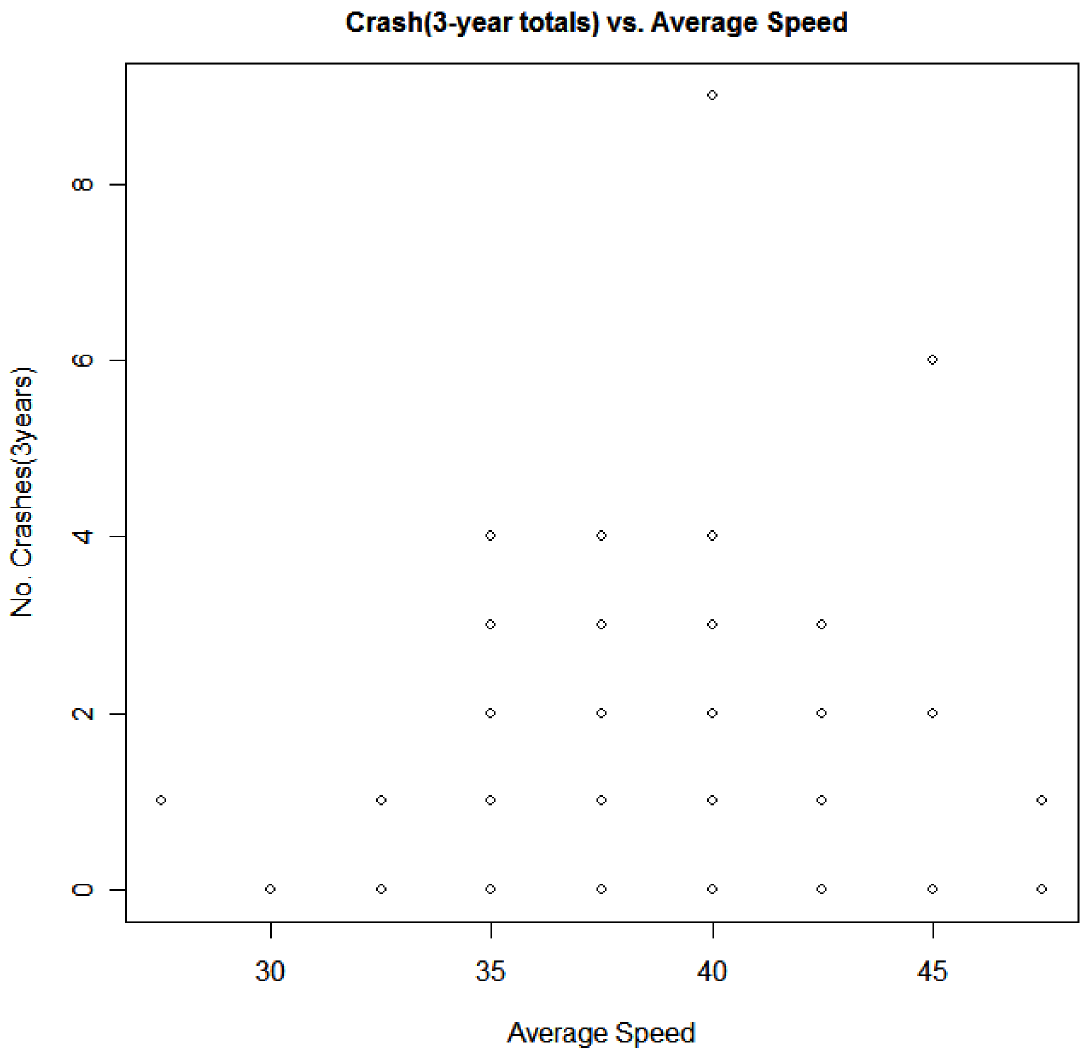

FIGURE 9 Crash distribution in terms of average speed

CHAPTER IV

METHODOLOGY

Model Description

Negative Binomial

There can be two definitions for the probability mass function of the negative binomial. One would be appeared in a statistical text book as follows:

If X_r represents a designated number of failures until r successes are achieved, then X_r is a negative binomial random variable with a parameter r (23). Bernoulii trials stop when r^{th} success is accomplished. The general negative binomial distribution has as the following form.

P(*x* failure before *r* successes) = P(*x* failures in *x +r-1* trials followed by a success)

= P(x failure in *x+r-1* trials) × P(*a* success) [by independence]

$$= \binom{x+r-1}{x} P^{r-1}(1-P)^x \times P$$

$$= \binom{x+r-1}{x} P^r (1-P)^x \tag{12}$$

Below equations show that this is a true probability distribution

$$\sum_{x=0}^{\infty} \binom{x+r-1}{x} P^r (1-P)^x = 1 \tag{13}$$

$$E(X_r) = \frac{r(1-P)}{P} \tag{14}$$

$$Var(X_r) = \frac{r(1-P)}{P^2} \tag{15}$$

This form, P(X_r=x)= $\binom{x+r-1}{x} P^r (1-P)^x$, is not often seen in the traffic safety area. The negative binomial used in a traffic safety area is a reparameterization of the original version of the negative binomial. Reparameterization is to express P in terms of μ and replace P with other expressions resulted from computation. Then, P is $\frac{\mu}{\mu+r}$ and 1-P is $\frac{r}{\mu+r}$. Again these equations can be substituted for P and 1-P. Then the probability mass function will be $P(X_x = x) = \binom{x+r-1}{x}\left(\frac{r}{\mu+r}\right)^r \left(\frac{\mu}{\mu+r}\right)^x$ where, r can be replaced by any notation such as θ, k, and α. If ϕ is used instead of r.

$$P(X_x = x) = \binom{x+\phi-1}{x}\left(\frac{\phi}{\mu+\phi}\right)^{\phi} \left(\frac{\mu}{\mu+\phi}\right)^x \tag{16}$$

This equation could be written using a factorial when ϕ is an integer.

$$P(X_x = x) = \frac{(x+\phi-1)!}{x!(\phi-1)!}\left(\frac{\phi}{\mu+\phi}\right)^{\phi} \left(\frac{\mu}{\mu+\phi}\right)^x \tag{17}$$

A binomial coefficient could be transformed in terms of a gamma function.

$$P(X_x = x) = \frac{\Gamma(x+\phi)}{x!\Gamma(\phi)}\left(\frac{\phi}{\mu+\phi}\right)^{\phi} \left(\frac{\mu}{\mu+\phi}\right)^x \tag{18}$$

This formula derived from the beginning is not very different from the original form. However, this formula has a gamma function, and ϕ could be any positive number. Thus, the meaning used in the Bernoulli trial before has been lost and ϕ is handled solely as a fitting parameter. A negative binomial distribution has two parameters: μ and ϕ

$$X \sim NegBin(\mu, \phi)$$

μ is an expected number of crashes at an intersection, and ϕ is a dispersion parameter.

The mean and variance can be rewritten as: *E(X)*= μ, *Var*(*X*)= $\dfrac{\mu(\mu+\phi)}{\phi^2} = \mu + \dfrac{\mu^2}{\phi}$

The next section will show in detail how to derive these equations. With the positive integer ϕ being used, the original interpretation of a negative binomial has been lost. A negative binomial should be considered as a distribution that can be flexibly fit discrete data instead.

Relationship between Poisson and Negative Binomial

The Poisson distribution is a limiting case of a Negative Binomial. Let ϕ tend to be infinity to see what happens on the negative binomial formula above. The formula derived above can be rewritten like below.

$$\frac{\Gamma(x+\phi)}{x!\Gamma(\phi)}\left(\frac{\phi}{\mu+\phi}\right)^r\left(\frac{\mu}{\mu+\phi}\right)^x = \frac{(x+\phi-1)!}{x!(\phi-1)!}\left(\frac{\phi}{\mu+\phi}\right)\left(\frac{\mu}{\mu+\phi}\right)^x$$

$$= \frac{(\phi+x-1)(\phi+x-2)....\phi(\phi-1)!}{x!(\phi-1)!}\left(\frac{\mu+\phi}{\phi}\right)^{-\phi}\frac{\mu^x}{(\mu+\phi)^x}$$

$$= \frac{(\phi+x-1)(\phi+x-2)....\phi}{(\phi+\mu)^x}\left(\frac{\mu+\phi}{\phi}\right)^{-\phi}\frac{\mu^x}{x!}$$

$$= \frac{\phi+x-1}{\phi+\mu}\cdot\frac{\phi+x-2}{\phi+\mu}\cdots\frac{\phi}{\phi+\mu}\left(\frac{\mu+\phi}{\phi}\right)^{-\phi}\frac{\mu^x}{x!}$$

If ϕ tend to be infinite, this formula will tend to be the Poisson formula.

$$\frac{\phi + x - 1}{\phi + \mu} \rightarrow 1$$

$$\frac{\mu^x}{x!} \rightarrow \frac{\mu^x}{x!}$$

$$\left(\frac{\mu + \phi}{\phi}\right)^{-\phi} \rightarrow e^{-\mu}\ \frac{\mu^x}{x!}$$

$$\lim_{\phi \to \infty} \frac{\Gamma(x+\phi)}{x!\Gamma(\phi)} \left(\frac{\phi}{\mu + \phi}\right)^{\phi} \left(\frac{\mu}{\mu + \phi}\right)^{x} \rightarrow \frac{e^{-\mu}\mu^x}{x!} \tag{19}$$

As the above equation has been derived, the Poisson distribution is the special case when ϕ goes to be infinite. This shows that the negative binomial can vary to different distributions as ϕ becomes altered. Here ϕ is utilized as a measurement data deviation which is Poisson distributed. If a dataset has a smaller ϕ value than the other one, it could be stated that the data are more clumped than the Poisson distribution typically does. For that reason, most commonly ϕ is called the dispersion parameter (or the over-dispersion parameter). With the two parameters defined above, the variance is stated in terms of ϕ and μ from the negative binomial.

$$Var(X) = \frac{r(1-p)}{p^2} = \frac{\phi \dfrac{\mu}{\mu + \phi}}{\left(\dfrac{\phi}{\mu + \phi}\right)^2} = \phi \frac{\mu}{\mu + \phi} \cdot \frac{(\mu + \phi)^2}{\phi^2}$$

$$= \quad \frac{\mu(\mu + \phi)}{\phi^2} = \mu + \frac{\mu^2}{\phi} \tag{20}$$

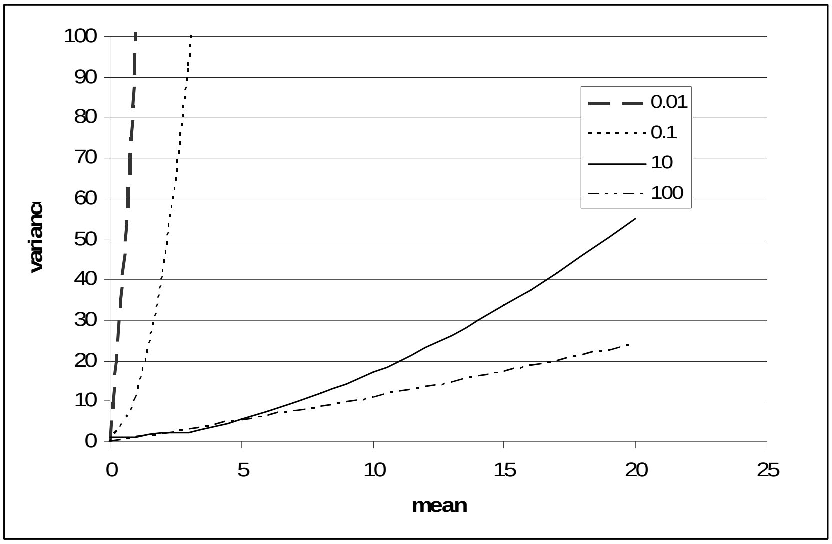

FIGURE 10 Negative binomial mean-variance relationship

Figure 10 shows that variance can vary in terms of the dispersion parameter ϕ. From that equation derived, as ϕ becomes infinite, Var(X) is getting closed to μ. That mean is not far from the variance which is typically seen in a Poisson distribution. The negative binomial can represent a whole range of heteroscedastic behavior with the corresponding ϕ value.

Model Development

Crashes are discrete, sporadic and independent events. Negative binomial distribution has been used to model the accident data in many past studies because it can overcome the problem of over-dispersion (5, 6, 7, 13). Additionally, it also can deal with temporal variability due to its extreme flexibility. For this reason, a negative binomial is employed to represent the accident data.

Due to the link function, the mean can be expressed as shown below:

$$\log(\mu) = \alpha + \beta_1 S_1 + \beta_2 S_2 + \cdots\cdots + \beta_n S_n \qquad (21)$$

The original model form will be like the following equation:

$$\mu = e^{\alpha} \cdot e^{S_1^{\beta_1} + S_2^{\beta_2} + S_3^{\beta_3} + \cdots\cdots + S_n^{\beta_n}} \tag{22}$$

Development of the models involves the determination of which explanatory variables should be employed, whether and how variables should be grouped, and how variables should be entered into models. Variables such as volume, control delay, V/C ratio, and speed were explored to identify the relevance between accident occurrence and traffic factors. A negative binomial was employed to do that on account of its advantages over Poisson. In a negative binomial regression, the coefficient estimation is computed by a maximum likelihood method. To do this computation, the following likelihood function will be used:

$$L(\mu,\phi) = \prod_{i=1}^{n} \frac{\Gamma(x_1+\phi)}{x_i!\Gamma(\phi)} \left(\frac{\phi}{\mu+\phi}\right)^{\phi} \left(\frac{\mu}{\mu+\phi}\right)^{x_i} \tag{23}$$

Take the logarithm of this equation, and calculate the coefficient that maximizes the likelihood function. All the coefficient values can be calculated through this process.

For example, if the model form is $\mu_i = \alpha S^{\beta_1} Y^{\beta_2}$, this can be expressed in terms of a negative binomial probability mass function. The function can be defined as:

$$L(\phi,\alpha,\beta_1,\beta_2;x,S,Y) = \prod_{i=1}^{n} \frac{\Gamma(x_1+\phi)}{x_i!\Gamma(\phi)} \left(\frac{\phi}{\alpha S^{\beta_1} Y^{\beta_2}+\phi}\right)^{\phi} \left(\frac{\alpha S^{\beta_1} Y^{\beta_2}}{\alpha S^{\beta_1} Y^{\beta_2}+\phi}\right)^{x_i} \tag{24}$$

After taking the log of their equation, the coefficient of ϕ, α, β_1 *and* β_2 can be estimated using statistical software R or SAS. In this study R is used.

Model Evaluation and Selection

Log-Likelihood Ratio Test

Likelihood ratio tests compare two models in order to determine if the reduced model is statistically different from the original model. A Log-likelihood test can be illustrated as a difference between log-likelihood of two models. The ratio of two likelihood functions is written as the following (13, 23, 24):

$$LRT = -2\ln\left(\frac{\zeta_s(\hat{\theta})}{\zeta_g(\hat{\theta})}\right) \tag{25}$$

L_s is the likelihood function for the reduced model. ζ_g is the likelihood function for the full model including all the parameters. The test statistic is Chi-square distributed with the degree of freedom equal to the difference in the number of parameters between the two models. It can be rewritten as:

$$\begin{aligned} LRT &= -2\left(\ln(\zeta_s) - \ln(\zeta_g)\right) \\ &= -2\ln(\zeta_s) + 2\ln(\zeta_g) \end{aligned}$$

$$= deviance_g - deviance_s \tag{26}$$

Thus, the LRT can be defined as a difference in the deviance for the two models. This is convenient as the deviance is a value of interest in other respects.

Akiake Information Criteria (AIC)

In the process of building a statistic model, a bias would be larger if too few variables were applied. Precision would be higher as number of variables increases because it is measured by the Standard Error estimation. The investigator should compromise between the precision and bias. Based on the fact that there is no true model illustrating the reality perfectly, the best model is to have the least loss of information. Kullback and Leibler (1951) illustrate a measure called as the Kullback-Leibler information when approximating the real world (25). Akiake has accomplished the relationship between a maximum-likelihood and Kullback-Leibler information. Therefore, AIC is the information criteria to estimate the Kullback-Leibler information. It can be expressed as:

$$\text{AIC} = -2(\log - \mathit{likelihood}) + 2k \qquad (27)$$

k is the number of parameters included in the model

The AIC value has no meaning in itself. It becomes meaningful when it is compared to the AIC of a series of models. The best model of all models specified is one with the lowest AIC value. The model selection requires the statistical modeler to consider the previous case, system characteristic, and the engineer's intuitive judgment. AIC will give the best of the poor models if only poor models are available. The AIC value also allows for a model comparison. It is a simple way to identify the difference between a candidate model and a best model. Generally, if the difference between two models is less than 2, the two models would not be considered to be much different (7)

CHAPTER V

EXPLORATORY DATA ANALYSIS

In this chapter, all the potential variables were explored to identify the relationship between perceived traffic factors as safety hazard and left-turn crashes. Negative binomial was employed to investigate the relationships. Not knowing directional information of crash, all the potential factors such as control delay, V/C ratio, and speed were calculated by average values except volume. Volume was taken log for convenience of computing process during modeling.

Left-Turn Type vs. Left-Turn Crashes

To quantify the effect of a left-turn type and compare them with one another, a regression model was conducted based on the negative binomial process. This modeling process provided clues of a relationship between a left-turn type and crashes over a targeted area. Table 8 basically has the same information as shown in Figure 11.

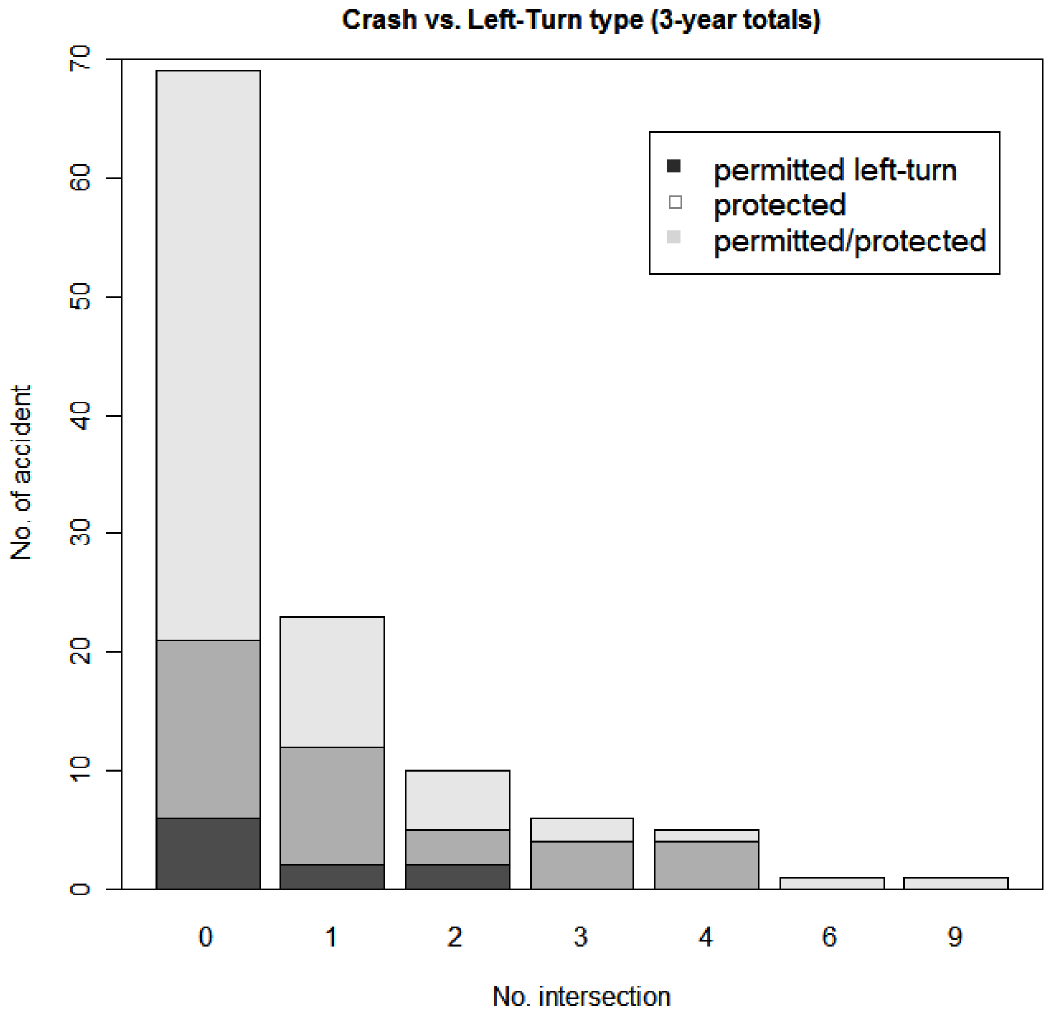

FIGURE 11 No. Crash vs. left-turn phasing type

TABLE 8 Left-turn phasing type

No. Crashes	0	1	2	3	4	6	9
Permitted	6	2	2	0	0	0	0
Perm/prot	15	10	3	4	4	0	0
protected	48	11	5	2	1	1	1

Table 9 indicates that if an intersection employs a permitted plus protected left-turn type, it would lead vehicles to more exposure to crashes than other left-turn types. On the other hand, if the intersections have permitted left-turns, it seems to have a less opportunity to be exposed to crashes. Permitted left-turns have a relatively small number of observations compared to two other left-turn types. Therefore, it is hard to tell the effect of a permitted left-turn on safety. Due to the lack of sample size, the coefficient estimate may be biased. Also, other variables such as volume or a control delay need to be considered to confirm that the effect of a left-turn type is consistent. Generally, intersections with permitted left-turns have a relatively low traffic volume. Low volume may not allow intersections to have a chance to have left-turn crashes as much as other left-turn types of intersections. Because of this, left-turn phasing type should be used along with other factors in developing a model to obtain more accurate results.

TABLE 9 regression result for left-turn phasing type

	Estimate	Std. Error	Z value	Pr(>|z|)
Intercept	-1.6094	0.5750	-2.799	0.00512 **
Permitted/protected	0.6574	0.6296	1.044	0.29643
Protected	0.1348	0.6138	0.220	0.82621
Permitted	NA	NA	NA	NA
AIC	292.88			
Dispersion parameter(standard error)	0.610 (0.189)			
Deviance	99.643			

Left-Turn Phasing Sequence vs. Left-Turn Crashes

There have been quite a few studies involving the effect of a left-turn scheme on safety performance at signalized intersections. A left-turn sequence should be considered as a significant factor of a left-turn type. In Table 10, coefficient estimates show that lead/lag seems to have the highest potential in left-turn crashes at signalized intersection. In order to confirm the presence of the yellow trap issue at signalized intersections, all the combinations of type and sequence were examined. Lead and protected combination is most widely used for a left-turn in the dataset. A left-turn type and sequence were applied for regression based on the negative binomial structure to quantify the effect of both factors. Figure 12 shows how left-turn crashes are distributed in terms of phasing sequence at signalized intersections over targeted area. Since only three observations employed lag sequences as phasing sequence, it was excluded from this box plot.

TABLE 10 regression result for left-turn phasing sequence

	Estimate	Std. Error	Z value	Pr(>|z|)
Intercept	-1.5386	0.2297	-6.697	2.13e-11 ***
Lead/Lag	0.7883	0.3837	2.054	0.0400 *
Split	0.1523	0.3952	0.385	0.7000
Lag	0.1523	0.8812	0.173	0.8628
Lead	NA	NA	NA	NA
AIC	293.28			
Dispersion parameter (standard error)	0.640(0.203)			
Deviance	100.10			

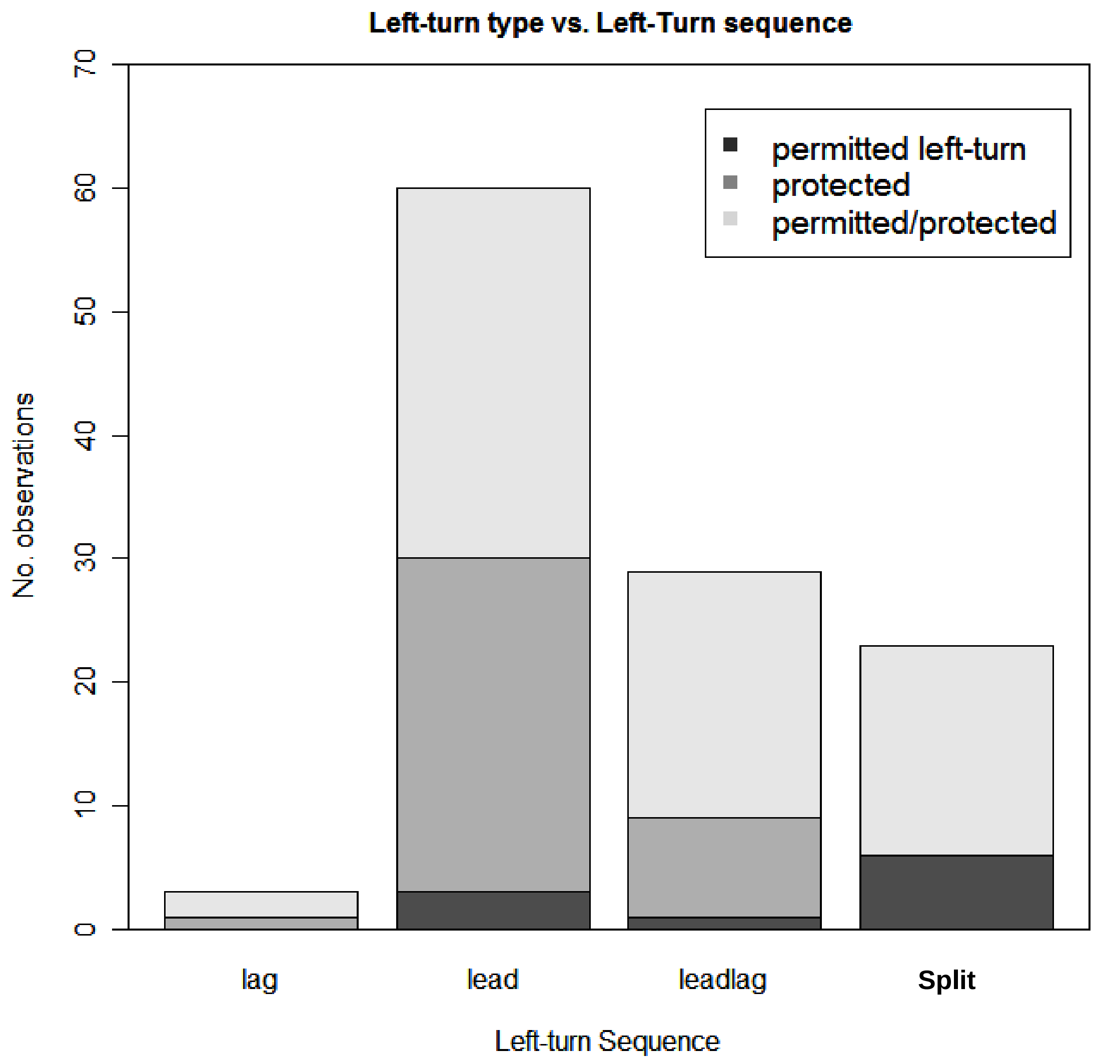

FIGURE 12 Phasing type and sequence

Tables 11, 12, and 13 indicate that when a left-turn type and sequence were used together; that is, a protected plus permitted left-turn and lead/lag sequence would contribute to an unsafe left-turn. The three regression models (Table 11, Table 12, and Table 13) illustrate the relationship between the crash count and the combination of a left-turn type and sequence. It seems that both the protected plus permitted type and lead-lag could have a left-turning driver face hazardous situations. Table 11 seems to be consistent with the fact that a lead-lag sequence with a protected/permitted left-turn lead left-turn drivers to face a yellow trap situation. Moreover, permitted left-turn types and lag phases appear to be the safest left-turn scheme. Nonetheless, other variables need to be investigated, since the sample size is very low and it may have correlation with other variables.

TABLE 11 Regression results for left-turn phasing sequence with permitted/protected left-turn

Variable	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-2.1909	0.3271	-6.699	2.11e-11 ***
Permitted/protected	1.0000	0.3607	2.772	0.00557 **
Lead/lag	1.2329	0.3900	3.162	0.00157 **
Split	0.6976	0.4441	1.571	0.11621
Lag	0.3481	0.8975	0.388	0.69813
Lead	NA	NA	NA	NA
AIC	287.89			
Dispersion parameter(standard error)	0.751(0.246)			
Deviance	99.182			

TABLE 12 Regression results for left-turn phasing sequence with protected left-turn

Variable	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-1.2864	0.2540	-5.065	4.09e-07 ***
Protected	-0.7797	0.3287	-2.372	0.01770 *
Lead/lag	1.1489	0.3926	2.926	0.00343 **
Split	0.3849	0.4036	0.954	0.34018
Lag	0.3213	0.8975	0.358	0.72037
Lead	NA	NA	NA	NA
AIC	290.02			
Dispersion parameter(standard error)	0.718 (0.233)			
Deviance	99.182			

TABLE 13 Regression results for left-turn phasing sequence with permitted left-turn

Variable	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-1.5320	0.2317	-6.613	3.76e-11 ***
Permitted	-0.2602	0.6309	-0.412	0.6800
Lead/lag	0.7817	0.3846	2.032	0.0421 *
Split	0.2161	0.4113	0.526	0.5992
Lag	0.1457	0.8810	0.165	0.8687
Lead	NA	NA	NA	NA
AIC	295.12			
Dispersion parameter(standard error)	0.642(0.203)			
Deviance	100.07			

Control Delay vs. Left-Turn Crashes

Earlier study illustrated the significance of a *level of service indicators* such as delay, V/C ratio, capacity etc. (18) in developing a safety prediction model in addition to volume data. A different traffic volume model is estimated for different traffic *level of service indicator* in their studies to improve the prediction accuracy. As a matter of fact, this study used average value for control delay since the given data do not indicate the directional information with respect to crashes. The relationship between control delay and number of crashes was plotted to identify the likelihood of a crash count as control delay varies. It seems that the control delay can be effective in changing the crash count. A negative binomial regression was applied with only the control delay as a variable to estimate the coefficient and validate its significance level. Figure 13 shows that congestion is associated with a higher risk of crashes. According to Table 14 and Table 15, the control delay that left-turn vehicles experience at signalized intersections has a significant effect on the number of crashes.

TABLE 14 Wald test for control delay

Response: Number of crashes			
Control delay	**Wald Chi-square**	**Degree of Freedom**	**Pr(>Chi-square)**
	11.465	1	0.0007094

TABLE 15 Regression result for control delay

	Estimate	**Std. Error**	**Z-value**	**Pr(|>Z|)**
Interceptor	-2.085408	0.262794	-7.936	2.10e-15 ***
Control Delay	0.013408	0.003556	3.770	0.000163 ***
AIC	283.21			
Dispersion parameter(standard error)	0.731(0.237)			
Deviance	99.448			
Functional Form	$\mu = 0.124256413 \cdot e^{(0.0134508 x_{control\_delay})}$ (28)			

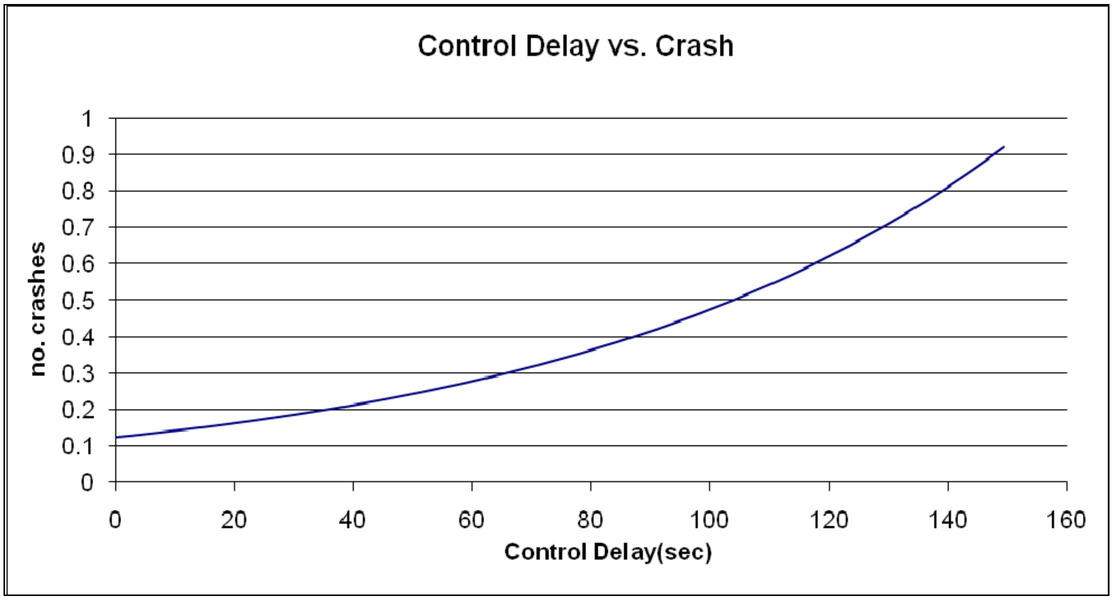

FIGURE 13 Plot expected number of left-turn crashes in terms of control delay

V/C ratio vs. Left-turn Crashes

The V/C ratio is the ratio of volume to be served to the road capacity. A V/C ratio was assumed to be another representative for a level of service in the current traffic situation. Data were plotted to examine the relationship between the V/C ratio and crashes at signalized intersections. From the plot, there seems to be a relationship between them but it is fairly obstreperous. A regression was applied for the dataset to investigate their relationships. As shown Figure 14, a higher V/C ratio seems to lead left-turn vehicles to encounter a higher left-turn crash risk at signalized intersections. The V/C ratio can be considered as a criterion to explain the current traffic state (i.e. level of service). Basically, it has similar functions to control delay. To quantify the effect of a V/C ratio, a coefficient was estimated based on a negative binomial structure. The Wald test indicates how the V/C ratio is related to the crash count. From both statistical methods, if V/C ratio is increasing, the vehicle would have a more chance to face the crash than the lower V/C ratio. Nonetheless, a V/C ratio range is from 0 to 2, and the coefficient is not very high even if it is statistically significant. Figure 14 shows that the fitted curve resulted from a regression based on a negative binomial distribution to identify such relationships between the two.

TABLE 16 Wald test for V/C ratio

Response: Number of crashes			
	Wald Chi-square	**Degree of Freedom**	**Pr(>Chi-square)**
V/C ratio	30.4599	1	3.408e-08 ***

TABLE 17 Regression result for V/C ratio

	Estimate	**Std. Error**	**Z-value**	**Pr(|>Z|)**
Intercept	-2.0454	0.3038	-6.733	1.67e-11 ***
V/C ratio	1.7136	0.5931	2.889	0.00387 **
AIC	285.64			
Dispersion parameter(standard error)	0.707(0.232)			
Deviance	100.50			
Functional Form	$\mu = 0.129328448 \cdot e^{(1.7136 x_{V/C\,ratio})}$ (29)			

As shown table 16 and 17, a higher V/C ratio seems to lead left-turn vehicles to encounter a higher left-turn crash risk at signalized intersections. The V/C ratio can be considered as a criterion to explain the current traffic state (i.e. level of service). Basically, it has a function similar to control delay. To quantify the effect of a V/C ratio, a coefficient was estimated based on a negative binomial structure. The Wald test indicates how the V/C ratio is related to the crash count. From both statistical methods, if the V/C ratio is increasing, the vehicle would have a more of a chance to face the crash than the lower V/C ratio.

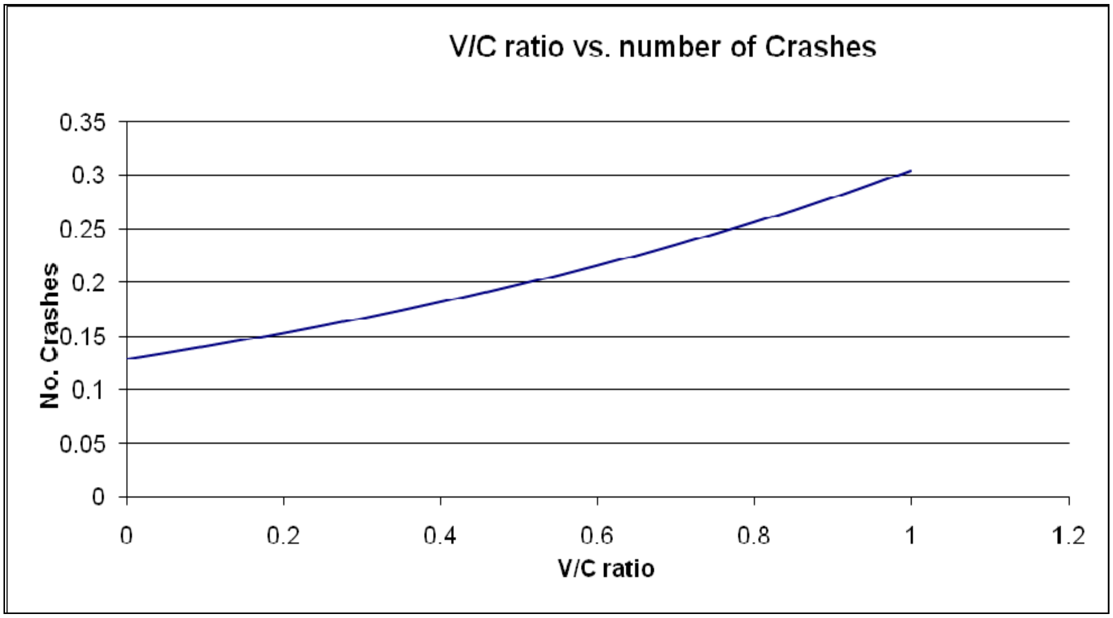

FIGURE 14 Plot expected number of left-turn crashes in terms of V/C ratio

Volume vs. Left-Turn Crashes

Left-Turning volume

Because a left-turn crash always involves left-turn vehicles, a left-turn approach volume was explored to quantify the effect of the left-turn volume on safety performance at intersections. As table 18 and 19 shown, volume is taken by the natural log because of convenience and following traditional way. The Wald test illustrates the significance of the log of left-turn volume as a variable for the prediction model. The regression line resulted from modeling is plotted in Figure 15. It clearly shows the relationship resulting from the regression. The model form should be different since volume is used as log form here.

TABLE 18 Wald test for left-turn volume

Response: Number of crashes			
	Wald Chi-square	**D.F**	**Pr(>Chi-square)**
Intercept	16.7473	1	4.27e-05 ***
Log_Left-Turn Volume	9.0088	1	0.002687 **

TABLE 19 Regression result for left-turn volume

	Estimate	**Std. Error**	**Z-value**	**Pr(|>Z|)**
Intercept	-4.6212	1.0687	-4.324	1.53e-05 ***
Log_left-turn volume	0.6239	0.1967	3.172	0.00152 **
AIC	284.50			
Dispersion parameter(ϕ)	0.702(0.226)			
Deviance	99.083			
Expected number of crashes (μ)	$\mu = 0.00984098 \cdot x_{left-turn\_volume}^{0.6239}$ (30)			

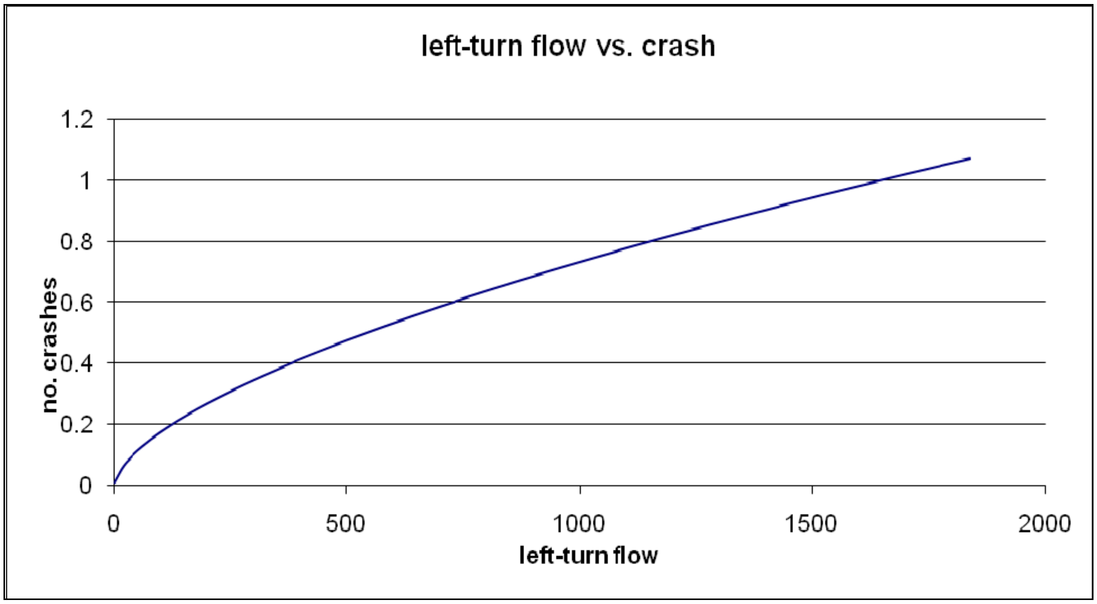

FIGURE 15 Plot expected number of left-turn crashes in terms of left-turn volume (vph)

Through Volume

Intersections with a permitted plus protected left-turn are likely to cause a left-turning vehicle to be involved in a crash with the opposing vehicles. It is easily perceived when a left-turning vehicle passes the middle of the intersection prior to the opposing vehicle passing that location on the permitted left-turn. More through volume pass the intersections, a left-turn vehicle would have more opportunities to encounter a crash risk with an opposing vehicle. With this insight, through volume is attempted to be used as a variable to identify the relationship between number of crashes and volume. Although the plot seems to be very noisy, the regression line shows a positive relationship between the number of crashes and volume. A non-parameter smooth seems to have a skewed quadratic form. To examine the relationship between them correctly, several statistic tests were conducted. As Tables 20 and 21 show, although coefficients for volume data are not significant for explaining much of variance according to the Wald test; it appears to contribute to the slight increasing of crash counts.

TABLE 20 Wald test for through volume

Response: Number of crashes			
	Wald Chi-square	**D.F**	**Pr(>Chi-square)**
Intercept	45.879 1	1	1.258e-11 ***
Log_Through Volume	1.519	1	0.2177

TABLE 21 Regression result for through volume

	Estimate	**Std. Error**	**Z-value**	**Pr(|>Z|)**
Intercept	-1.4749651	0.2300296	-6.412	1.44e-10 ***
Log_Through volume	0.08336	0.0002159	1.167	0.243
AIC	292.8			
Dispersion parameter(standard error)	0.580(0.176)			
Deviance	99.363			
Expected number of left-turn crashes	$\mu = 0.00984098 \cdot x_{through\_volume}^{0.08336}$			(31)

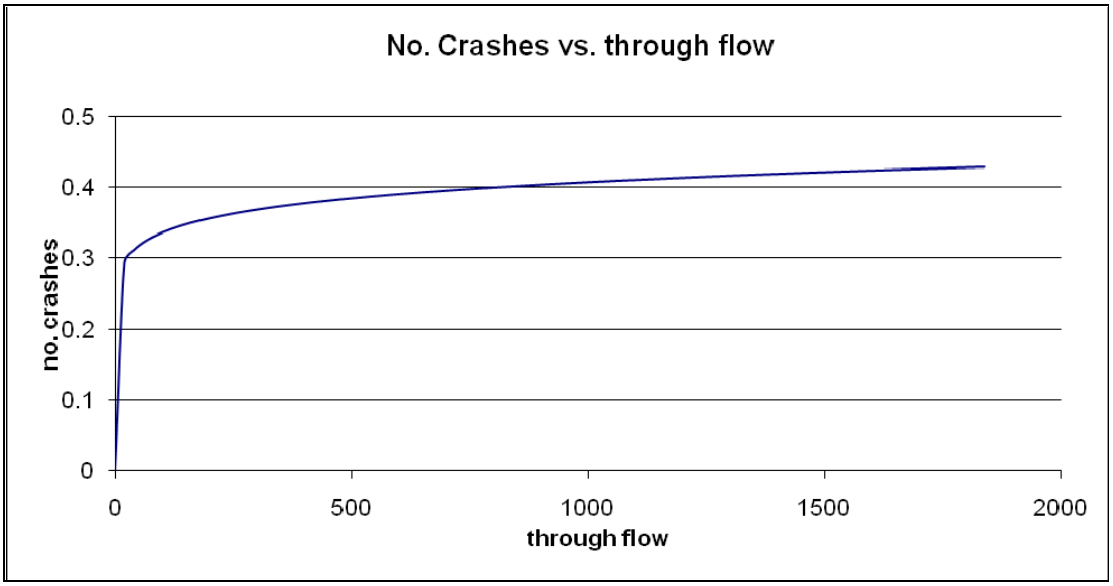

FIGURE 16 Plot expected number of left-turn crashes in terms of through volume (vph)

Intersecting Road Volume

Of all left-turn crash types, quite a few angle collisions are observed in the collected data. It suggests that intersecting volume may be a factor. Based on the intuition, the relationship between the crash and intersecting road volume was investigated. As in the other volume factors shown beforehand, Table 22 shows some trends for left-turn crashes associated with the intersecting road volume. Table 22 and Table 23 show the results of the prediction model for the intersecting volume–crash relationships. The coefficient of intersecting volume is significant at the 95% level. To explicitly examine the relationship for crash and intersecting volume, plots were produced as for the other volumes. Similar to through volume, there is not as much change on the crash count as volume changes.

TABLE 22 Wald test for intersection volume

Response: Number of crashes			
	Wald Chi-square	**D.F**	**Pr(>Chi-square)**
Intercept	50.885	1	9.794e-13 ***
Log_Intersecting Volume	15.172	1	9.814e-05 ***

TABLE 23 Regression result for intersecting volume

	Estimate	**Std. Error**	**Z-value**	**Pr(|>Z|)**
Intercept	-8.9287	1.9591	-4.558	5.18e-06 ***
Log_Intersecting volume	1.0239	0.2593	3.949	7.85e-05 ***
AIC	277.11			
Dispersion parameter(standard error)	0.819(0.274)			
Deviance	97.689			
Expected number of left-turn crashes (μ)	$\mu = 0.00013253 \cdot x_{intersecting\_volume}^{1.0239}$ (32)			

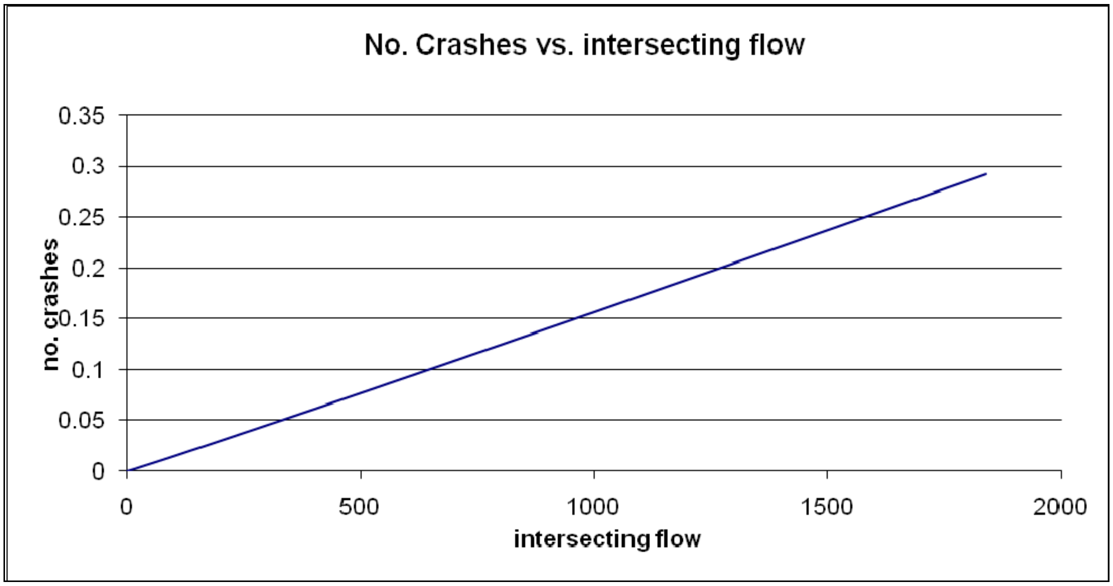

FIGURE 17 Plot expected number of crashes in terms of intersection volume (vph)

Intersecting Volume and Through Volume Combined

In an attempt to strengthen the relationship between the volume and crash count, through and intersecting volume were combined. In table 24 and 25, the coefficient of the combined volume is statistically significant at the 95% level. The plot is produced based on the result in Table 25. Figure 18 seems to show that more explicit relationships are observed. It shows that combined volume appears to be more clear-cut than when the volumes were used individually. Since the coefficient of combined volume is larger than any types of volume used beforehand, the fitted curve has concave shape.

TABLE 24 Wald test for combined volume

Response: Number of crashes			
	Wald Chi-square	**D.F**	**Pr(>Chi-square)**
Intercept	43.225	1	4.88e-11 ***
Log_combined volume	15.015	1	0.0001067 ***

TABLE 25 Regression result for combined volume

	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-12.2131	3.1882	-3.831	0.000128 ***
Log_combined volume	1.3910	0.4052	3.433	0.000597 ***
AIC	280.58			
Dispersion parameter(standard error)	0.852(0.307)			
Deviance	102.66			
Expected number of left-turn crashes (μ)	$\mu = 0.000004965 \cdot x_{combined\_volume}^{1.3910}$ (33)			

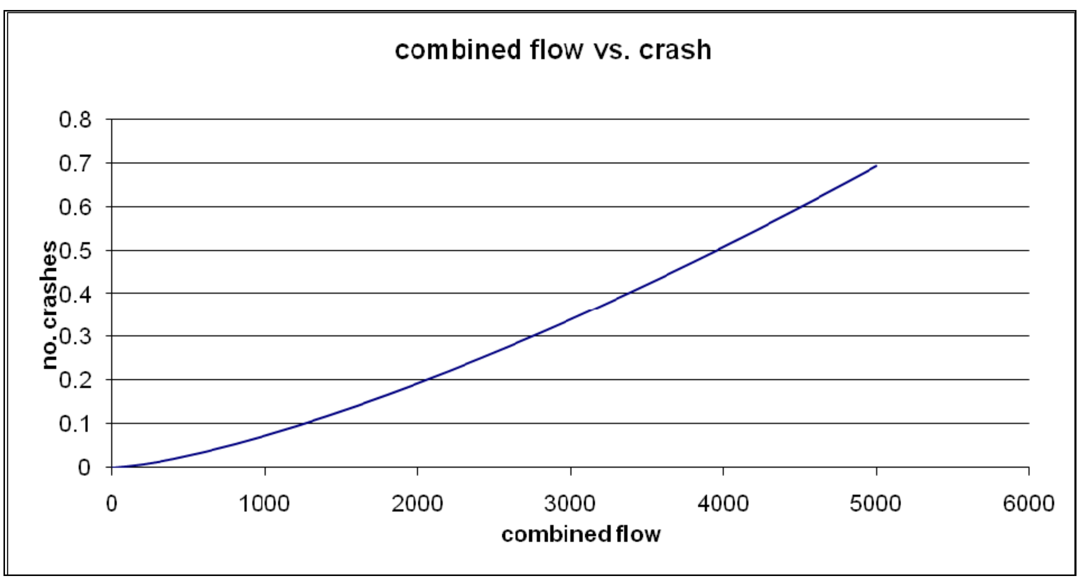

FIGURE 18 Plot expected number of crashes in terms of combined volume (vph)

Speed vs. Left-Turn Crashes

Knowing the considerable effect of speeds, average speeds (primary and intersecting roads) were used as a variable to identify the relationship between speed and number of crashes. The plot of speed against number of left-turn crashes was produced to examine the relationship. Figure 19 shows that the numbers of crashes seem to increase as vehicle speed increases. This indicates that a vehicle at higher speed has a higher likelihood to be involved in a crash than at lower speed. In Table 26 and 27, a coefficient of speed is significant at the 95% level. The plot produced based on Table 27 also identifies the number of crashes increased with speed.

TABLE 26 Wald test for speed

Response: Number of crashes			
	Wald Chi-square	**D.F**	**Pr(>Chi-square)**
Intercept	4.5439	1	0.03304 *
Average_Speed	4.0639	1	0.04381 *

TABLE 27 Regression result for speed

	Estimate	**Std. Error**	**Z-value**	**Pr(|>Z|)**
Intercept	-4.14895	1.54482	-2.686	0.00724 **
Average_Speed	0.07455	0.03993	1.867	0.06192 .
AIC	290.61			
Dispersion parameter(standard error)	0.608(0.187)			
Deviance	99.252			
Expected number of left-turn crashes (μ)	$\mu = 0.015781 \cdot e^{0.07455 x_{average\_speed}}$ (34)			

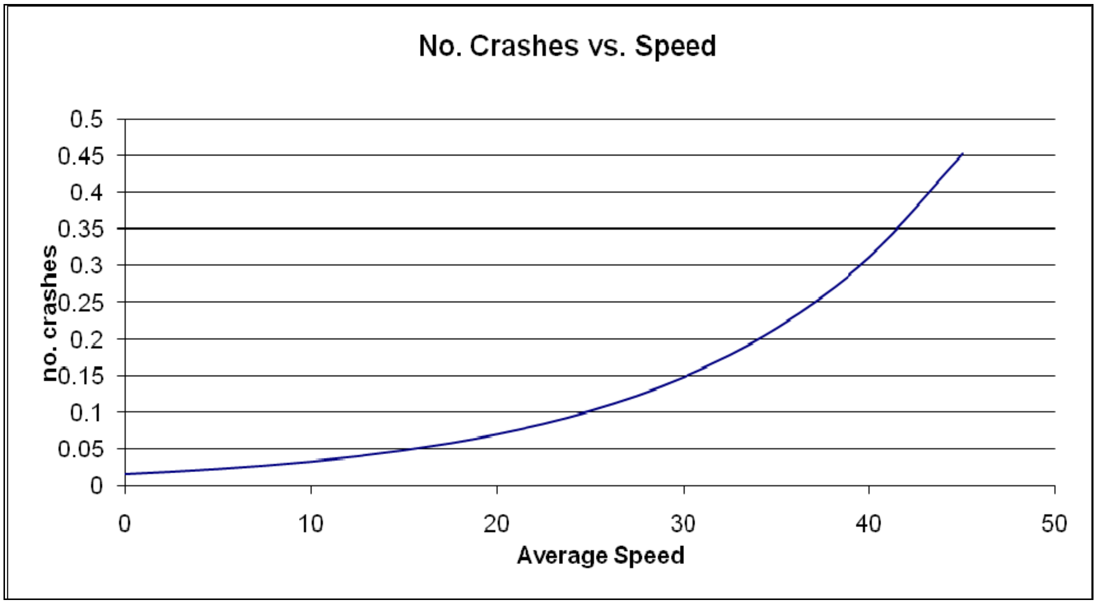

FIGURE 19 plot expected crashes in terms of average speed (mph)

As presented before, several potential factors were identified through the exploratory data analysis. All the factors except for phasing type, sequence, and through volume are identified as potential factors affecting the change in left-turn safety at signalized intersections at the 95% level of confidence. However, since many have correlation between them and more variables can produce better fit for the observed data, multiple regressions need to be developed to enhance prediction accuracy for future safety performance prediction.

CHAPTER VI

ANALYSIS AND RESULTS

Process of Model Selection

In Chapter V, explanatory variables have been explored to identify how each factor is related to the crash counts individually. The plots produced beforehand provided a preliminary understanding. To identify the number of left–turn crashes associated with traffic operations, a multivariate model needs to be employed. Generally, a predictive model with more explanatory variables is better to describe the crash count than one with fewer variables. To select the model having the best fit, various tools such as deviance, Akiake Information Criteria (AIC), and log-likelihood ratio can be used. However, using those tools as a model selection may not be the best way for choosing the best model. The model selection does not depend only on those tools but also on a logical sense. A variable selection should also be according to an engineer's intuitive understanding on the traffic operations and experiences for searching the best model. Furthermore, it is assumed that all the variables are to be statistically independent with other variables in multivariate analysis. Multicollinearity conflicts with basic assumption of multivariate analysis. It implies that possible correlations between predictor variables are significant. To see whether this problem exists in given data, correlation analysis is conducted. As Table 28 shown, all the variables are kept since there is no remarkable correlation.

In Table 29 (model I), it seems that all the variables except permitted left-turns have a positive effect on increasing the left-turn crash. As for type and sequence, a protected left-turn and lag phase are reference variables respectively. Coefficients for the variables (i.e. different type of left-turn and phase) have no meaning as a value. Those values become meaningful when compared to each other in the same category. A lag phase is relatively the safest left-turn scheme compared to other three different phases: lead, split, and lead/lag. Likewise, a permitted left-turn has a relatively lower safety risk compared to other two left-turn types: protected and permitted plus protected. Generally, the intersection with permitted left-turns is deemed to have a higher safety risk. The results from Table 29(model I) might be counterintuitive. On the other hand, a permitted left-turn is implemented at the intersection that has fewer left-turn vehicles. As quite a few previous studies have shown that crash counts occurring at the intersection are accounted for approaching traffic volume, permitted left-turn could be safer not because of safer left-turn type but because of less volume. Also there are not many intersections with permitted only left-turns as most of the intersections are majors ones on the arterials with significant left-turn and through volume. Volume, speed, and permitted/protected left-turn types appear to have a considerable contribution to increase left-turn crashes at the 95% level of significance. Although lead/lag left-turn phases do not reach the 95% level of significance by a slight difference (approximately 0.0107), it seems to be able to account for a left-turn crash. It seems that the lag phase has a better safety performance at the intersection. However, this case is similar to a permitted left-turn case in that there are not many intersections with a lag phase. As the matter of fact, only three intersections have a lag left phase out of all observations.

TABLE 28 Results of the correlation analysis

	Left-turn volume	Combined volume	Lead/lag	split	lag	Permitted/ protected	Permitted	Control Delay	V/C Ratio
Left-turn Volume									
Combined Volume	-0.22								
Lead/lag	-0.21	0.05							
Split	0.01	0.07	0.39						
Lag	-0.24	-0.12	0.23	0.15					
Permitted/ Protected	-0.05	0.16	0.30	0.44	0.09				
Permitted	0.01	-0.10	0.12	-0.23	0.05	0.10			
Control Delay	-0.26	-0.02	0.02	-0.08	0.02	-0.19	0.07		
V/C ratio	-0.21	-0.18	-0.04	0.10	0.07	0.10	-0.19	0.07	
Speed	-0.26	0.26	0.10	-0.11	-0.06	-0.08	0.09	0.34	-0.08

TABLE 29 Summary of model I

Variables	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-18.435821	3.854083	-4.783	1.72e-06 ***
Left-turn volume	0.074560	0.221027	0.337	0.735865
Combined volume	1.349456	0.393621	3.428	0.000607 ***
Control delay	0.003273	0.002088	1.568	0.116992
V/C ratio	0.125755	0.332918	0.378	0.705626
Lead	0.564268	0.810892	0.696	0.486516
Lead/Lag	1.480919	0.807776	1.833	0.066754
Split	1.269484	0.847263	1.498	0.134046
Lag	NA	NA	NA	NA
Permitted/protected	0.883031	0.321137	2.750	0.005965 **
Permitted	-0.045615	0.556601	-0.082	0.934684
Protected	NA	NA	NA	NA
Posted Speed limit	0.056246	0.019219	2.927	0.003427 **
AIC	271.66			
Dispersion parameter (standard error)	2.06(1.05)			
deviance	102.1			
Functional Form	$\mu = \alpha \cdot F_{LT}^{s_1} \cdot F_{comb}^{s_2} \cdot e^{\sum_{i=1}^{n} \beta_i x_i}$			(35)

Table 30 (model II) presents the relationship between volumes (left-turn and combined) and number of crashes. In exploratory data analysis, it is speculated that volume variations can describe crash variations. The given results indicate that the number of crashes increase with volume. From the coefficient, both types of volume have positive values for the coefficient at the 95% level of significance. The sum of the intersecting road volume and through volume has a larger influence on the number of crashes than left-turn approach volumes. It appears that models with only volume variables are not accurate as much as the previous model with most variables. Although volume is the most usable variable in explaining variations of crash counts, it cannot describe the true nature of an accident (2). Besides volume, speed was also significant to explain the variation of a left-turn crash count in previous investigations. It is evident that volume and speed could better describe future safety performance at intersections. Table 31 shows the case that speed is added to model. Model III (Table 31) has better fit than Model II (Table 30) in terms of AIC and dispersion parameter.

TABLE 30 Summary of model II

Variables	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-12.6884	3.0915	-4.104	4.06e-05 ***
Left-turn volume	0.4642	0.2068	2.244	0.02483 *
Combined volume	1.1356	0.4263	2.664	0.00773 **
AIC	277.42			
Dispersion parameter(standard error)	0.89(0.311)			
deviance	99.081			
Functional Form	$\mu = 0.000003087 \cdot F_{LT}^{0.4642} \cdot F_{comb}^{1.1356}$ (36)			

TABLE 31 Summary of model III

Variables	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-17.39515	3.76713	-4.618	3.88e-06 ***
Left-turn volume	0.36571	0.20578	1.777	0.07553
Combined volume	1.35391	0.43046	3.145	0.00166 **
Posted Speed limit	0.09136	0.03964	2.305	0.02119 *
AIC	274.26			
Dispersion parameter(standard error)	1.058 (0.399)			
Deviance	99.67			
Functional Form	$\mu = 0.000000042 \cdot F_{LT}^{0.36571} \cdot F_{comb}^{1.35391} \cdot e^{(0.09136 X_{speed})}$ (37)			

The model should be considered to take either a V/C ratio or a Control delay since they represent the similar characteristic of intersections. Table 32 and Table 33 show a difference of two models in terms of using identical variables except one variable (V/C ratio and control delay). As Table 32 and Table 33 shows, both models seem to have identical results. Although they appear to be similar, the control delay would be preferred to use as a variable for a better prediction of left-turn crashes. The coefficient of the control delay has a better explanation of crash counts on account of the coefficient level of significance and the AIC.

Table 34 added the phase sequence as a variable to Table 32 model for identifying the effect of a left-turn phasing scheme. A left-turn scheme has been investigated by many researchers for its safety concern at an intersection. Compared to lead phase, lead/lag has a relatively worse degree of safety for a left-turn and lag phase seems to be the best phasing scheme. Only the lead/lag phase had a relatively significant coefficient among four different left-turn plans even though it is not at the 95% level of significance. The lead/lag is easily seen along with the permitted plus protected left-turn for causing yellow trap issues. The left-turn type is used along with the phase as variables in an attempt to investigate this combination.

TABLE 32 Summary of model IV

Variables	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-16.727687	3.673256	-4.554	5.27e-06 ***
Left-turn volume	0.082789	0.222587	0.372	0.70994
Combined volume	1.177366	0.419996	2.803	0.00506 **
Control delay	0.010729	0.004092	2.622	0.00875 **
Speed	0.132173	0.040869	3.234	0.00122 **
AIC	271.5			
Dispersion parameter(standard error)	1.262(0.528)			
Deviance	100.84			
Functional Form	$\mu = 0.000000054 \cdot F_{LT}^{0.082789} \cdot F_{comb}^{1.177366} \cdot e^{(0.04285 X_{speed} + 0.010729 X_{control\_delay})}$ (38)			

TABLE 33 Summary of model V

Variables	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-16.69775	3.90119	-4.280	1.87e-05 ***
Left-turn volume	0.22057	0.22517	0.980	0.32732
Combined volume	1.25949	0.44490	2.831	0.00464 **
V/C ratio	0.98832	0.66239	1.492	0.13568
Speed	0.10112	0.03999	2.529	0.01144 *
AIC	273.9			
Dispersion parameter(standard error)	1.086(0.405)			
Deviance	98.456			
Functional Form	$\mu = 0.000000095 \cdot F_{LT}^{0.22057} \cdot F_{comb}^{1.25949} \cdot e^{(0.10112 X_{speed} + 0.98832 X_{V/C\_ratio})}$ (39)			

The resulting model (Model VI) in Table 34 seems to strongly support the presence of safety problem associated with the yellow trap in the targeted area. The yellow trap issue is easily observed when a permitted plus protected left-turn and lead/lag phasing scheme are used together at the intersection. Adding a phasing type as a variable to table 34, table 35 shows that the effect of a lead/lag on the left-turn crash count becomes more significant than when a left-turn phase is used alone. Coefficients of other variables are not very different from the previous model.

TABLE 34 Summary of model VI

Variables	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-16.980866	3.599088	-4.718	2.38e-06 ***
Log_Left-turn volume	0.096022	0.223614	0.429	0.667626
Log_Combined volume	1.168716	0.396331	2.949	0.003190 **
Control delay	0.010557	0.003867	2.730	0.006329 **
Lead	NA	NA	NA	NA
Lead/Lag	0.526725	0.335710	1.569	0.116650
Split	0.207551	0.349275	0.594	0.552355
Lag	-0.759833	0.812190	-0.936	0.349512
Posted Speed limit	0.134492	0.040462	3.324	0.000888 ***
AIC	272.6			
Dispersion parameter(standard error)	1.569(0.726)			
Deviance	102.22			
Functional Form	$\mu = 0.000000042 \cdot F_{LT}^{0.096022} \cdot F_{comb}^{1.168716} \cdot e^{(0.134492 X_{speed} + 0.010557 X_{contor\_delay})}$ (40)			

TABLE 35 Summary of model VII

Variables	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-18.284227	3.573958	-5.116	3.12e-07 ***
Log_Left-turn volume	0.072590	0.217946	0.333	0.739087
Log_Combined volume	1.365483	0.387308	3.526	0.000423 ***
Lead	NA	NA	NA	NA
Lead/Lag	0.843883	0.341456	2.471	0.013458 *
Split	0.633199	0.394082	1.607	0.108105
Lag	-0.602006	0.807438	-0.746	0.455924
Permitted/protected	0.807603	0.316871	2.549	0.010813 *
Permitted	-0.014476	0.553404	-0.026	0.979131
Protected	NA	NA	NA	NA
Posted Speed limit	0.121768	0.040654	2.995	0.002743 **
Control Delay	0.008205	0.003728	2.201	0.027761 *
AIC	269.81			
Dispersion parameter(standard error)	2.09(1.09)			
deviance	102.67			
Functional Form	$\mu = 0.000000011 \cdot F_{LT}^{0.072590} \cdot F_{comb}^{1.365483} \cdot e^{(0.008205 X_{control\_delay} - 0.602006 X_{lag} + 0.843883 X_{lead/lag} + 0.633199 X_{split} + 0.807603 X_{per/prot} - 0.014476 X_{per} + 0.121768 X_{speed})}$ (4			

Table 36 shows a predictive model (Model VIII) employs all variables that are perceived to be affecting left-turn crashes. Since lead can be a used as the reference variable and lag is rarely seen in this area, the lead/lag is used as a variable. Split is not significant at 95% level of confidence. As for a left-turn type, a permitted/protected left-turn is used as the only variable. The AIC is by far the lowest one and every coefficient is significant at the 95% level except the left-turn. Table 37 (model IX) uses very similar variables and has statistical results to what model VIII employed except split and V/C ratio instead of control delay. However, as table 37 shown, V/C ratio and split phasing is not significant at 95% level of confidence.

TABLE 36 Summary of model VIII

Variables	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-17.195471	3.512141	-4.896	9.78e-07 ***
Log_Left-turn volume	0.014391	0.209089	0.069	0.94513
Log_Combined volume	1.262662	0.388444	3.251	0.00115 **
Control delay	0.008805	0.003852	2.286	0.02226 *
Lead/Lag	0.647055	0.305648	2.117	0.03426 *
Permitted/protected	0.577552	0.281850	2.049	0.04045 *
Posted Speed limit	0.128798	0.040701	3.165	0.00155 **
AIC	268.19			
Dispersion parameter(standard error)	1.851 (0.940)			
Deviance	104.10			
Expected number of crashes per site	$\mu = 0.000000011 \cdot F_{LT}^{0.014391} \cdot F_{comb}^{1.262662} \cdot e^{(0.008805 X_{control\_delay} + 0.647055 X_{lead/lag} + 0.577552 X_{per/prot} + 0.128798 X_{speed})}$ (42)			

TABLE 37 Summary of model IX

Variables	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-17.66026	3.70363	-4.768	1.86e-06 ***
Log_left-turn volume	0.10959	0.21184	0.517	0.60492
Log_combined volume	1.39253	0.40186	3.465	0.00053 ***
V/C ratio	0.84099	0.59832	1.406	0.15985
Lead/Lag	0.94301	0.33987	2.775	0.012298 *
Split	0.68481	0.39434	1.737	0.00553.
Permitted/protected	0.96502	0.31998	3.016	0.00256 **
Posted Speed limit	0.09464	0.03989	2.372	0.01767 *
AIC	269.06			
Dispersion parameter(standard error)	1.730(0.810)			
Deviance	101.27			
Expected number of crashes per site	$\mu = 0.000000011 \cdot F_{LT}^{0.10959} \cdot F_{comb}^{1.39253} \cdot e^{\substack{(0.84099 X_{V/C\_ratio} + 0.94301 X_{lead/lag} + 0.68481 X_{split} \\ + 0.96502 X_{per/prot} + 0.09464 X_{speed})}}$ (43			

TABLE 38 Summary of model X

Variables	Estimate	Std. Error	Z-value	Pr(|>Z|)
Intercept	-17.755984	3.484130	-5.096	3.46e-07 ***
Log_combined volume	1.335996	0.362830	3.682	0.000231 ***
Control Delay	0.008741	0.003391	2.577	0.009953 **
Lead/Lag	0.916080	0.325884	2.811	0.004938 **
Split	0.682947	0.378186	1.806	0.070942
Permitted/protected	0.830989	0.314789	2.640	0.008295 **
Posted Speed limit	0.121743	0.038598	3.154	0.001610 **
AIC	265.02			
Dispersion parameter(ϕ)	2.05(1.08)			
Deviance	103.42			
Expected number of crashes per site	$\mu = 0.000000019 \cdot F_{comb}^{1.335996} \cdot e^{(0.008741 X_{control\_delay} + 0.916080 X_{lead/lag} + 0.682947 X_{split} + 0.830989 X_{per/prot} + 0.121743 X_{speed})}$ (44)			

If the model selection depends only on the AIC value, then the left-turn volume would be removed from the variables used in the model. However, a left-turn volume should be used because a left-turn crash cannot occur without a left-turn approach volume. As tables(from 29 to 38) presented before, out of nine candidates only two appear to be close to the best model in terms of AIC, the log-likelihood ratio, deviance, and logical sense. The difference between the two models is whether the use of a V/C ratio or control delay as a variable for representing a level of the service indicator. In addition, model VIII does not include split phasing since it is not significant at 95% level of confidence. Both models were appraised whether they are feasible to describe crash characteristics on both logical and quantitative sides. Based on the modeling assessment, model VIII seems to be more appropriate to explain the number of crash since control delay has a larger effect on the variation of crash counts in terms of coefficients and the variable's range. Furthermore, the control delay has a coefficient at the 95% level of confidence while the V/C ratio is not significant. Despite this result, it is hard to insist that Model VIII is better for forecasting future safety performances at a signalized intersection. Figure 27 and Figure 28 are almost identical in shape. In table 38, model X seems to be more feasible to describe left-turn crash trend than the previous two models. However, model X does not include left-turn volume. As presented beforehand, left-turn volume is the main source for occurring left-turn crashes. If left-turn volume does not exist at intersection, left-turn crashes would never occur. Thus, model X is excluded from the candidates for best model even if it has very similar statistics to the previous two. Appendix A has that all the plots show the relationship between predicted values from models and observed values from field data. Graphs provide lowess line and linear regression line to identify models' fit explicitly. "lowess line" is used to explore the relationship between dependent variables and response variables without fitting a specific model. LOWESS stands for LOcally-WEighted Scatter plot Smoother(25).

Recommended Left-Turn Crash Model

Models VIII and IX would be recommended to use for modeling a left-turn crash data due to the goodness-of-fit and the model's logic. The model form can be expressed as:

$$\mu = \alpha F_1^{\beta_0} F_2^{\beta_1} e^{(S_1 X_1 + S_2 X_2 + S_3 X_3 + S_4 X_4 + S_5 X_5)} \tag{45}$$

α= interceptor

F_1= left-turn volume (vph)

F_2= volume combined (vph)

β_0= coefficient for left-turn volume

β_1= coefficient for volume combined

S_1= coefficient for V/C ratio(model IX) and Control Delay (model VIII)

S_5= coefficient for speed

X_1= V/C ratio (model IX) and control delay (model VIII)

X_2= lead/lag sequence phase

X_3= split phase

X_4= permitted/protected type

X_5= speed (mph)

S_2= coefficient for lead/lag sequence phase

S_3= coefficient for split

S_4= coefficient for permitted/protected

TABLE 39 Model comparison

	Model VIII	Model IX
A	-17.195471	-17.66026
β_0	0.014391	0.10959
β_1	1.262662	1.39253
S_1	0.008805	0.84099
S_2	0.647055	0.94301
S_3	N/A	0.68481
S_4	0.577552	0.96502
S_5	0.128798	0.09464
AIC	268.19	269.06
Deviance	104.10	101.27
Dispersion parameter (standard error)	1.851 (0.940)	1.730(0.810)

According to statistical criteria such as AIC, deviance, and dispersion parameter, both models seem to be very similar as the in Table 39. As mentioned before, factors considered are almost same expect for split phasing. A few different things have been found from the results. First, model VIII shows that the impact of the left-turning approach volume on a left-turn crash is relatively low compared to the combined volume.

On the other hand, coefficient($\boldsymbol{\beta_0}$) for left-turn volume and coefficient($\boldsymbol{\beta_1}$)for combined volume are not different as much as model VIII. The combined volume can explain the variation of an expected crash better than the left-turn volume; whereas, both volumes do not have much different impact on a left-turn crash. In addition to discrepancy of volume coefficients, the control delay has more impact on number of left-turn crashes than the V/C ratio. This is because the control delay values are from 6 to 150 and V/C values are from 0.2 to 1. If the values within this range are plugged in the equation based on the assumption that other conditions are identical, it can be easily observed that control delay yields a higher chance of a crash. Control delay is significant for the change of a crash count because of the nonlinear relationship between control delay and v/c ratio. There might be a relationship between the combined volume and the level of service indicator since volume should be included in computing those traffic parameters. From the results, lead/lag phasing sequence potentially leads the vehicle to unsafe situations according to the coefficient. The intersection with lead/lag sequence has approximately 2.7 times as many chances as other intersections with lead left or split phasing. A permitted plus protected left-turn phasing leads a vehicle to have a higher potential for an accident.

CHAPTER VII

CONCLUSIONS, LIMITATIONS, AND FUTURE WORK

Many studies have shown that the intersection is the main safety concern in a traffic system due to the right-of-way issue for many conflicting movements. Different types of left-turns and left-turn sequences are often implemented to enhance the intersection operation efficiency. Studies have also been conducted using statistics model for estimating the left-turn safety performance at an intersection in terms of intersection characteristics (5, 13). Some of the developed models mainly focused on identifying the relationship between a crash and the implementation of different types of a left-turn and sequence.

In the exploratory data analysis, it is speculated that many traffic elements have a relationship with the crash count. They are not quite different from variables used in previous studies. It implies that there might be several factors which can be applied for describing the crash count under any condition and location. In this study, such factors as volume and level of service indicators were further investigated.

This study evaluated the impact of different characteristics of intersections on left-turn crashes at signalized intersections in College Station, TX. In addition to volume and level of service indicator, left-turn crashes occurring at intersections from 2002 to 2004 were evaluated. The left-turn crash count for different types of left-turns: protected, permitted, and permitted plus protected and sequence: lead/lag, lag, split, and lead were calculated. Negative binomial regressions were used to analyze the left-turn crash count by various intersection characteristics including left-turn phasing and sequence.

Volume is a significant factor to explain the variation of a left-turn crash count. However, a left-turn volume is not very significant for the change in safety according to modeling results. Additionally, it is noticed that contributions of a left-turn volume may vary depending on which variable is used as a level of service indicators.

Because of the yellow trap issue, an intersection having a permitted/protected left-turn phasing with a lead/lag left-turn is more likely to cause crashes. According to the results of this study, a lead/lag left-turn sequence is not significant at the 95% level of confidence without a permitted plus protected left-turn type in the developed model. This shows that the effect of a lead/lag left-turn sequence is significant only with a permitted plus protected left-turn type and vice versa. This seems to indicate that there could be a presence of a yellow trap problem and it might lead vehicles to have exposure to higher potentials of vehicle crashes. In this study, model VIII indicates that an intersection with a yellow trap problem has approximately 6.5 more times the chance of left-turn crashes.

Overall, a permitted plus protected left-turn phasing is the highest crash likelihood of all three left-turn types. The lead/lag sequence has the most impact on a left-turn safety at an intersection. Speed is able to explain the change in safety of a signalized intersection as what one would expect. Generally speaking, driver could not be able to have enough time to commit an appropriate maneuver, especially for old drivers. Furthermore, drivers might have more difficult situation when they need to make a left-turn at a high speed.

Conclusively, volumes seem to have a similar role of change in the safety at signalized intersections as identified in previous studies. More congestion is likely to lead left-turn vehicles to safety hazard situations similar to the previous study (11). As for speed, it contributes to crashes at intersections as well. Protect plus permitted left-turn combining with lead/lag left-turns seems to have the highest crash risk. Therefore, advanced signal type such as Dallas phasing is recommended to install at the location where yellow trap problems may exist to reduce the number of left-turn crashes. Additionally post speed limit should be considered to be lowered for the intersection with frequent crash experience without yellow trap problems.

Findings

The following findings are from the results and analysis. The results describe the impact of various traffic factors on left-turn crashes.

- Traffic operational characteristics can have an impact on left-turn safety.

- Intersecting volume is one of the most significant factors leading to left-turn crashes.
- A left-turn volume is very good at explaining the trend of a left-turn crash in the model which has solely volume as a variable.
- As opposed to what one would expect, when other factors were involved in a model as variables, coefficient for left-turn volume became not significant at the 95% level of confidence The level of service indicator can be used for explaining a crash trend.
- Intersecting road volume is significant for describing the crash trend.
- Combined volume is much better to explain the change in safety compared to using an individual volume.
- As speed increases, the number of left-turn crashes increase at the 95% level of confidence.

- Left-turn signal operations affect left-turn safety
  - Left-turn phasing types and phasing sequences are likely to affect the left-turn crash count.
  - Left-turn phasing types should be used along with a phasing sequence to identify the combined effect of the phasing type and sequence on safety at intersections.
  - Permitted plus protected combined with lead/lag left-turns have a significance in increasing the left-turn crash count even if they are not very significant individually for influencing a left-turn crash count.
- The model selection should be depending on the logical sense as well as statistical criteria.

Limitations

The study is limited by a number of constraints. The limitations listed below may have affected the result.

- The number of crashes at signalized intersections is limited.
- There is no indication on exact accident locations at the intersection (only street name is available).
- All traffic elements such as the volume, control delay, and V/C ratio were calculated by averaging values for two directions, as the data does not provide directional information.
- The data collection was conducted during peak period. Generally, there is a large difference by the directions during peak periods. If so, the results would be biased.
- If the volume has a large variance, an average volume is not appropriate to develop the prediction model.

Future Work

First of all, sufficient data should be collected to develop the model with accurate prediction ability. This is generally problematic for the prediction models developed in traffic safety areas if there are not sufficient data available. More data can produce more accurate prediction models for intersection safety. Alternatively to avoid the model being biased, advanced statistical methods such as the Bayesian method can be employed. However, the Bayesian method could produce a biased result for safety performance at signalized intersections without accurate prior estimations. The study develops a model for forecasting future left-turn crashes and estimates the contribution factors. If all the data are correct and the modeling process does not have any faults, it is probably possible to estimate the accident cost based on model predictions of crash occurrence. Implementations of lead/lag left-turn phasing sequences are based on operation considerations. It can save the green time which could have been otherwise consumed inefficiently by the direction with much smaller volume. If value of time can be estimated, it might be possible to determine which option is a more economical decision between a lead/lag left-turn phasing sequence and other left-turn phasing sequences. This can be done by choosing the most economic alternatives based on cost comparison between the accident cost and time saving benefits.

REFERENCES

1. Roess, R. P., E. S. Prassas., W. R. Mcshane. *Traffic Engineering third edition*. Peason Education, Inc. 2004. Upper Saddle River, NJ, 2004.

2. Lord, D., A. Manar, and A. Vizioli. "Modeling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments." Accident Analysis & Prevention. Vol. 37, N0.37, Amsterdam, NL, 2005, pp.185-199.

3. George, S. Accident data in College Station (2001~2003), College Station Police Department. College Station, TX. 2006.

4. Transportation Research Board. *Highway Capacity Manual 2000*. U.S. Department of Transportation, Washington, D.C., 2000.

5. Kulmala, R. "Safety at Rural Three- and Four-Arm Junctions." Development and Application of Accident Prediction Models. *Technical Research Center at Finland*, VTT Publications, Espoo, FIN. 1995.

6. Miaou, S.P. "The relationship between truck accidents and geometric design of road section: *Poisson versus negative binomial regression*." Accident Accident Analysis & Prevention Vol. 26, No. 4, Amsterdam, NL,1994, pp. 471-482.

7. Shankar, V.N., F. Mannering, and W. Barfield. Effect of roadway geometric and environmental factors on rural freeway accident frequencies. Accident Analysis & Prevention Vol. 27, No. 3, 1995, pp. 371~389.

8. Noyce, D. A., D. B. Fambro, and K. C. Kacir. Traffic Characteristics of Protected/Permitted Left-Turn Signal Displays. *Transportation Research Record*, TRB, National Research Council, Washington, D.C. January, 2000, pp. 28-39.

9. Hauer, E., Ng, JCN and Lovell, J. Estimation of safety at Signalized intersections. *Transportation Research Record*. TRB, National Research Council, Washington, D.C.,1988, pp. 48-61.

10. Retting, A., F. Chapline, and A. F., Williams. "Changes in crash risk following re-timing for traffic signal change intervals." Accident Analysis and Prevention Vol. 34. No. 2, 2002, Amsterdam, NL, pp. 210-220.

11. Morocoima-Black, R., S. Chavarria, and H. Kang. Selected Intersection Crash Analysis for 1996-2000. Champaign County Regional Planning Commission, Il, 2003.

12. Lyon, C., A. Haq, B. Persaud, and S. Kodama. "*Development of Safety Performance Functions for Signalized Intersections in a Large Urban Area and Application to Evaluation of Left-Turn Priority Treatment."* Presented at the Transportation Research Board 84th Annual meeting. Washington D.C., 2005.

13. Bauer, K. and D. Harwood. *Statistical Models of At-Grade Intersections-Addendum.* Report NO. FHWA-RD-99-094. Federal Highway Administration, Washington, D.C., 2000.

14. Chin, H. C. and M. A. Quddus. "Applying the random effect negative binomial models to examine traffic accident occurrence at signalized intersection." Accident Analysis and Prevention. Vol.35 No. 2, Amsterdam, NL, 2003, pp.253-259.

15. Mc Gee, H., S. Taori, and B. Persaud. *NCHRP Report 491: Crash Experience Warrant for Traffic Signals.* National Coorperative Highway Research Program, TRB, Washington, D.C., 2003.

16. Abdel-Aty, M. and J. Keller. "Exploring the overall and specific crash severity levels at signalized intersections." Accident Analysis and Prevention, Amsterdam, NL,Vol. 37, 2005, pp. 417-425.

17. Do-Gyeong Kim and S. Washington. "The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models." Accident Analysis & Prevention. Vo. 38, No. 6, Amsterdam, NL, 2006, pp. 1094-1110.

18. Burchett, G. D. and T. H. Maze. Rural Express Intersection Characteristics that Contribute to Reduce Safety Performance. *Mid-Continent Transportation Research Symposium*, Ames, Iowa, 2005.

19. Persaud, B. and T. Nguyen. *Safety Considerations in Capacity Analysis.* Department of Civil Engineering, Ryerson Polytechnic University, Toronto, Canada. 2005.

20. Chan, Ching-Yao. *Observation of Left-turn Conflicts at Signalized intersection and design of Gap-assistance system in Urban Environment.* Presented at the Transportation Research Board 85th Annual meeting. Washington D.C., 2006.

21. Persaud, B. and T. Nguyen. Disaggregate safety performance models for signalized intersections on Ontario provincial roads. *Transportation Research Record 1635,* TRB, National Research Council, Washington, D.C. 2006, pp.113-120.

22. Hallmark, S. L. and K. Mueller. *Impact of Left-turn Phasing on Older and Younger Drivers at High Seep Signalized Intersections.* CTRE Project 03-149 Center of Transportation Research and Education, Iowa State University, IA, 2004.

23. Lindsey, J.K.. *Applying Generalized Linear Models*. Springer-Verlag. New York, 1997.

24. Fox, J. *An R And S-Plus Companion To Applied Regression*. Sage Publications, Inc. London, UK, 2002.

25. Kullback, S., and R. A. Leibler. On information and sufficiency. Annals of Mathematical Statistics. Vol. 22, 1951, pp. 79-86.

APPENDIX A

Figures show the relationship between expected number of crashes and observed number of crashes. From the results, when the model has relatively good fit according to statistical criteria, lowess line seems to have similar shape to linear regression.

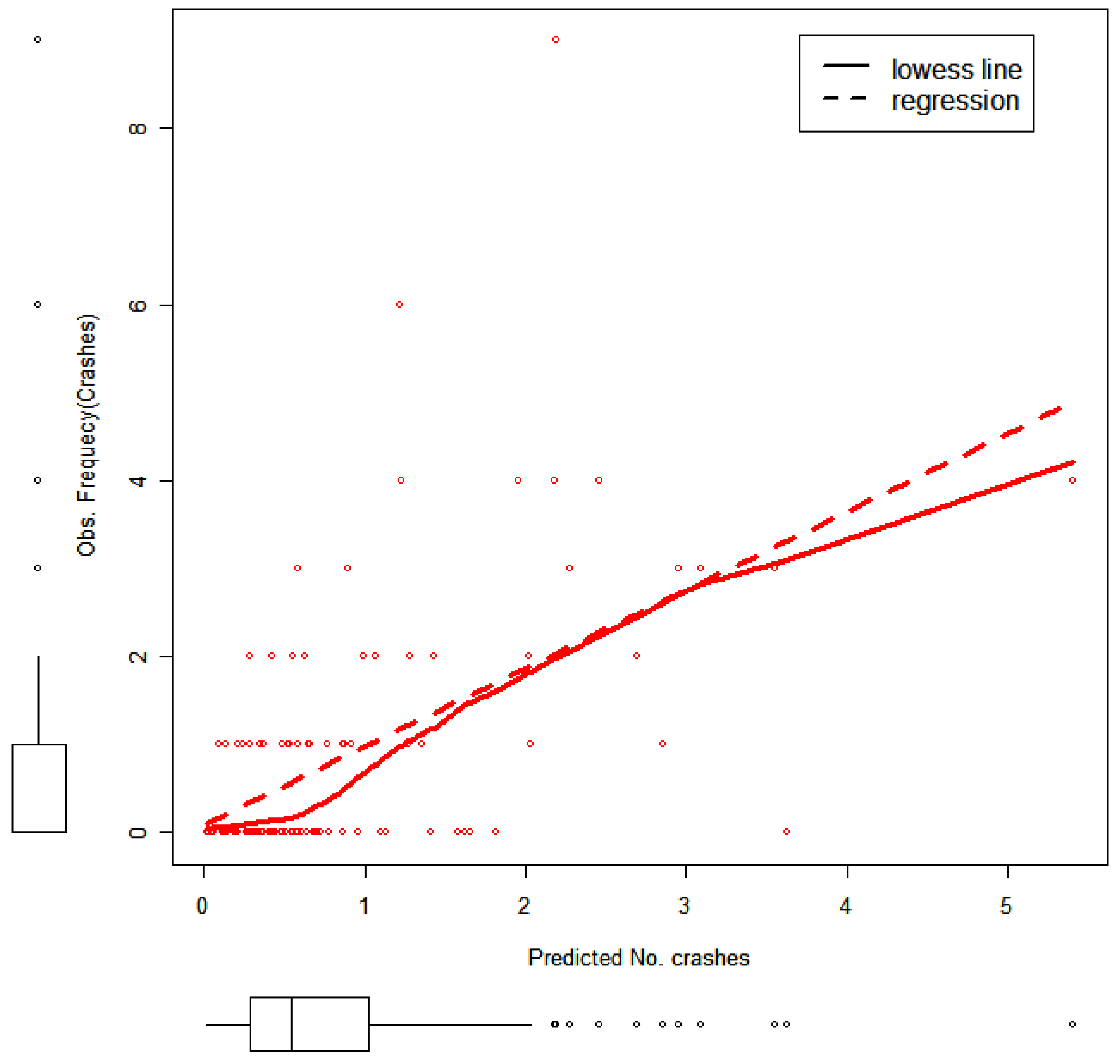

FIGURE 20 Model I

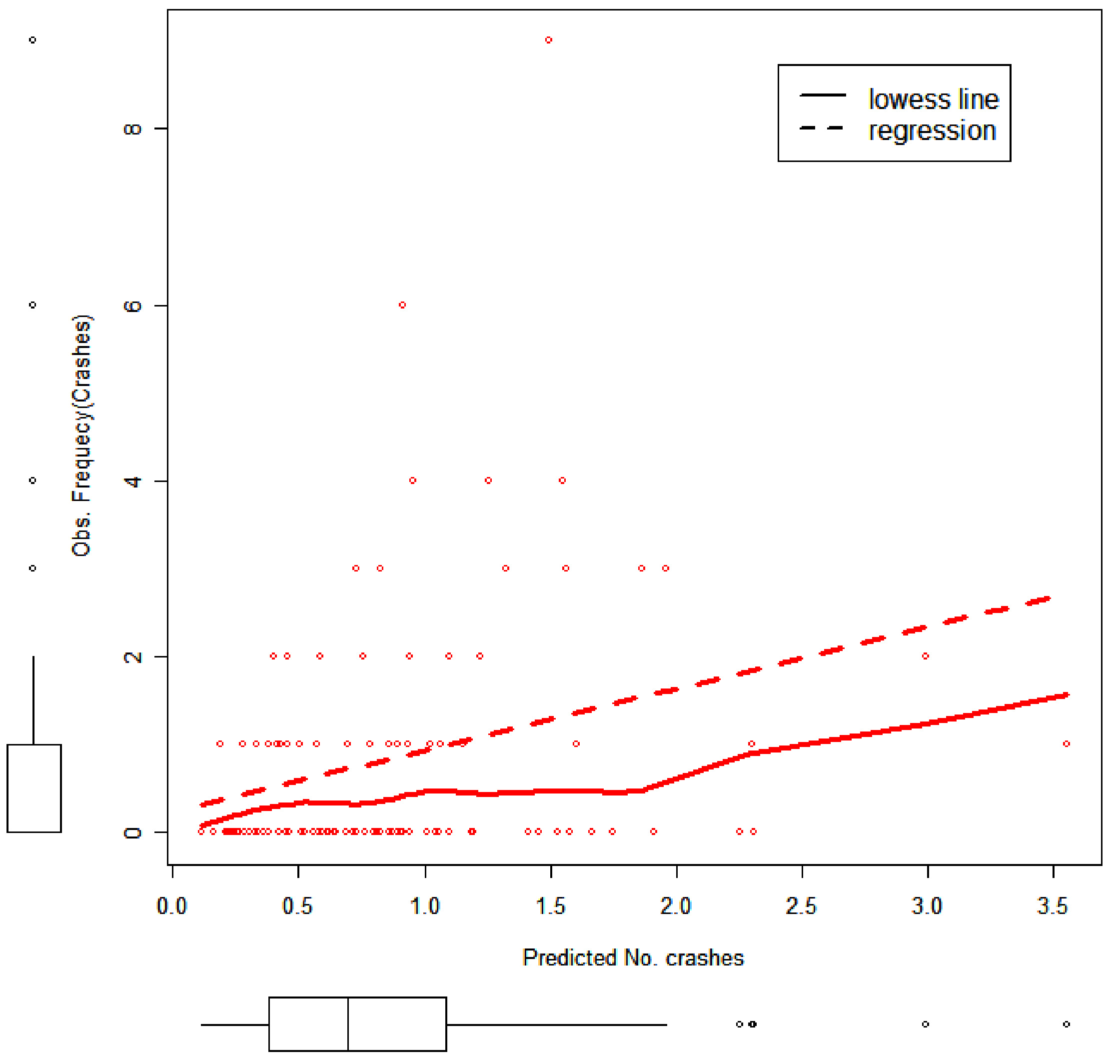

FIGURE 21 Model II

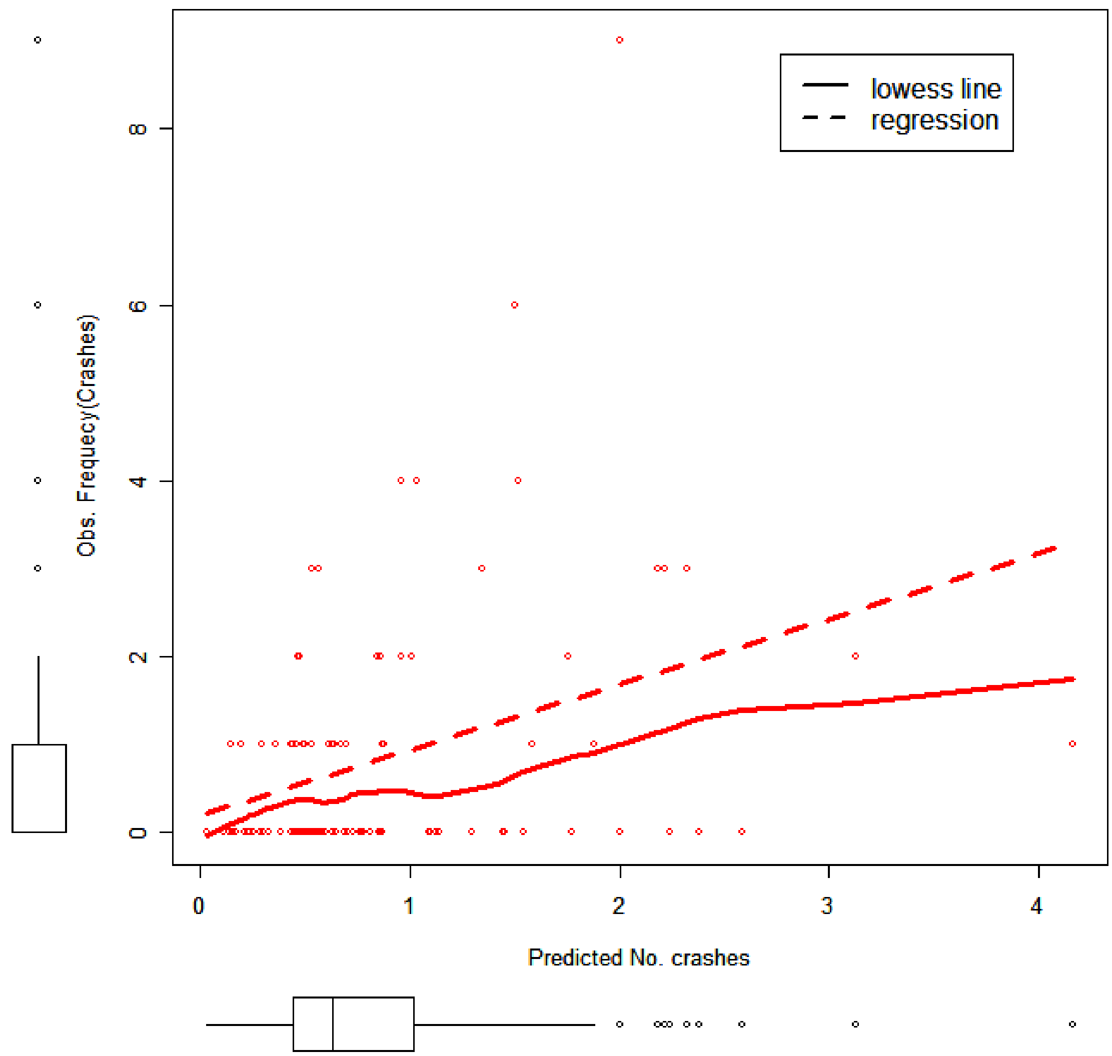

FIGURE 22 Model III

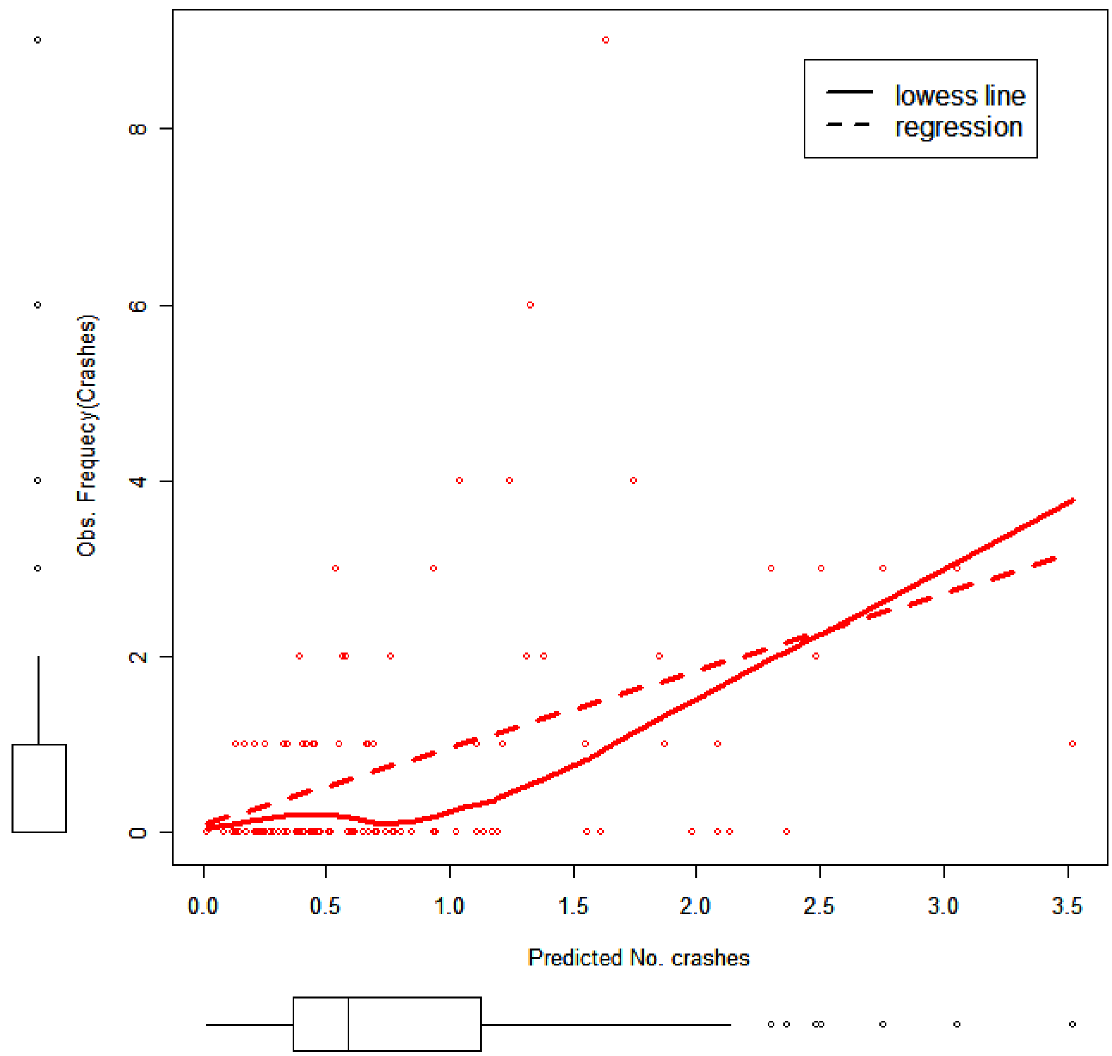

FIGURE 23 Model IV

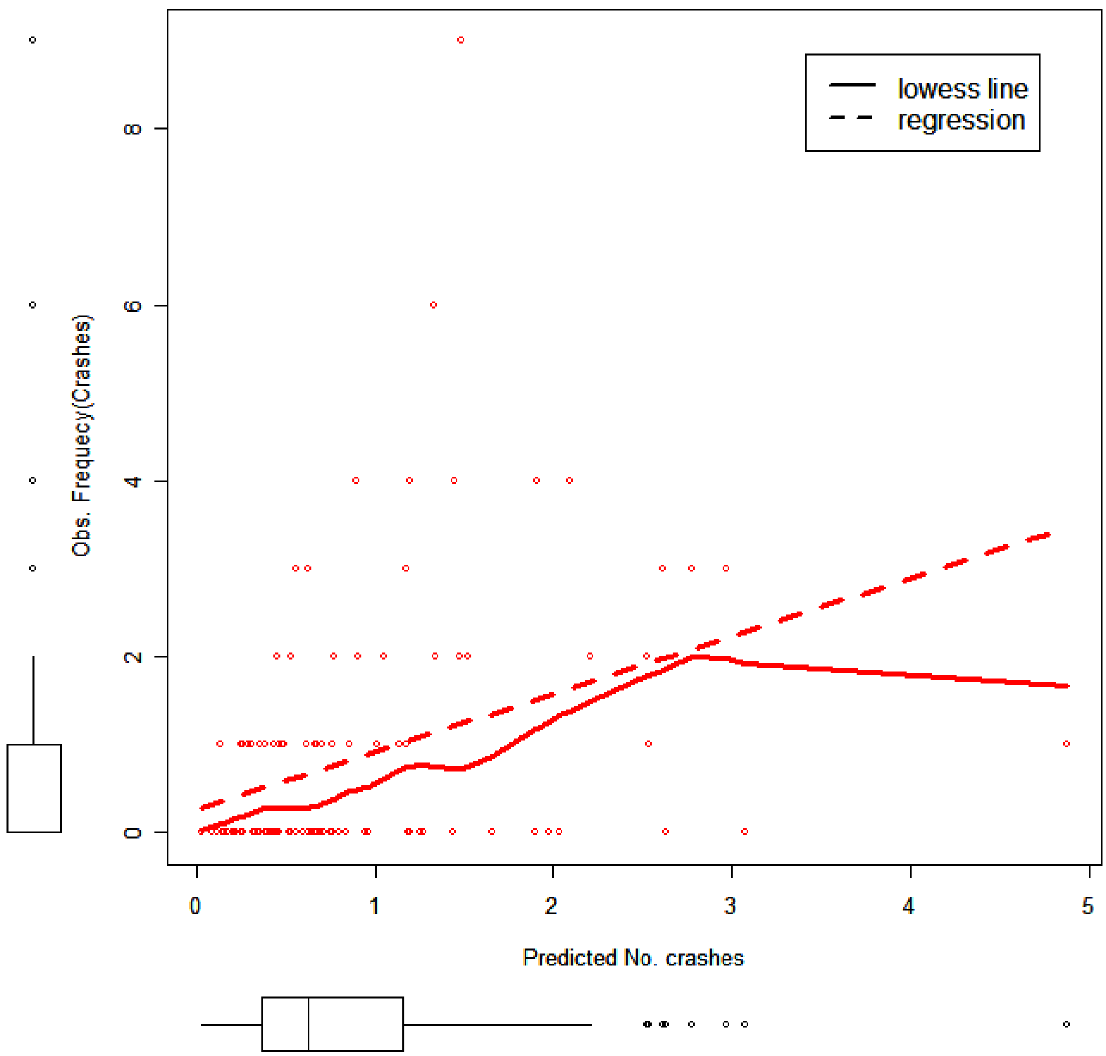

FIGURE 24 Model V

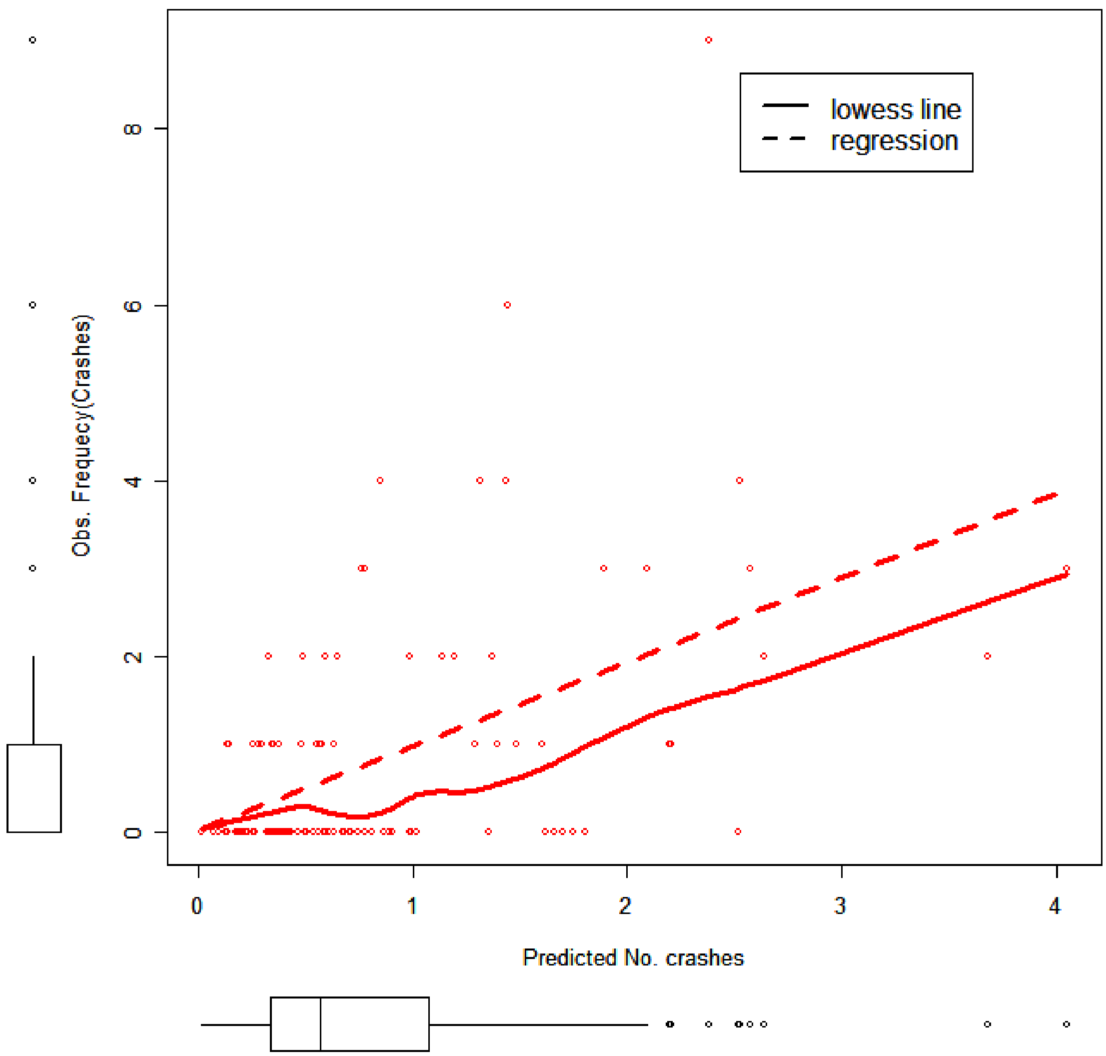

FIGURE 25 Model VI

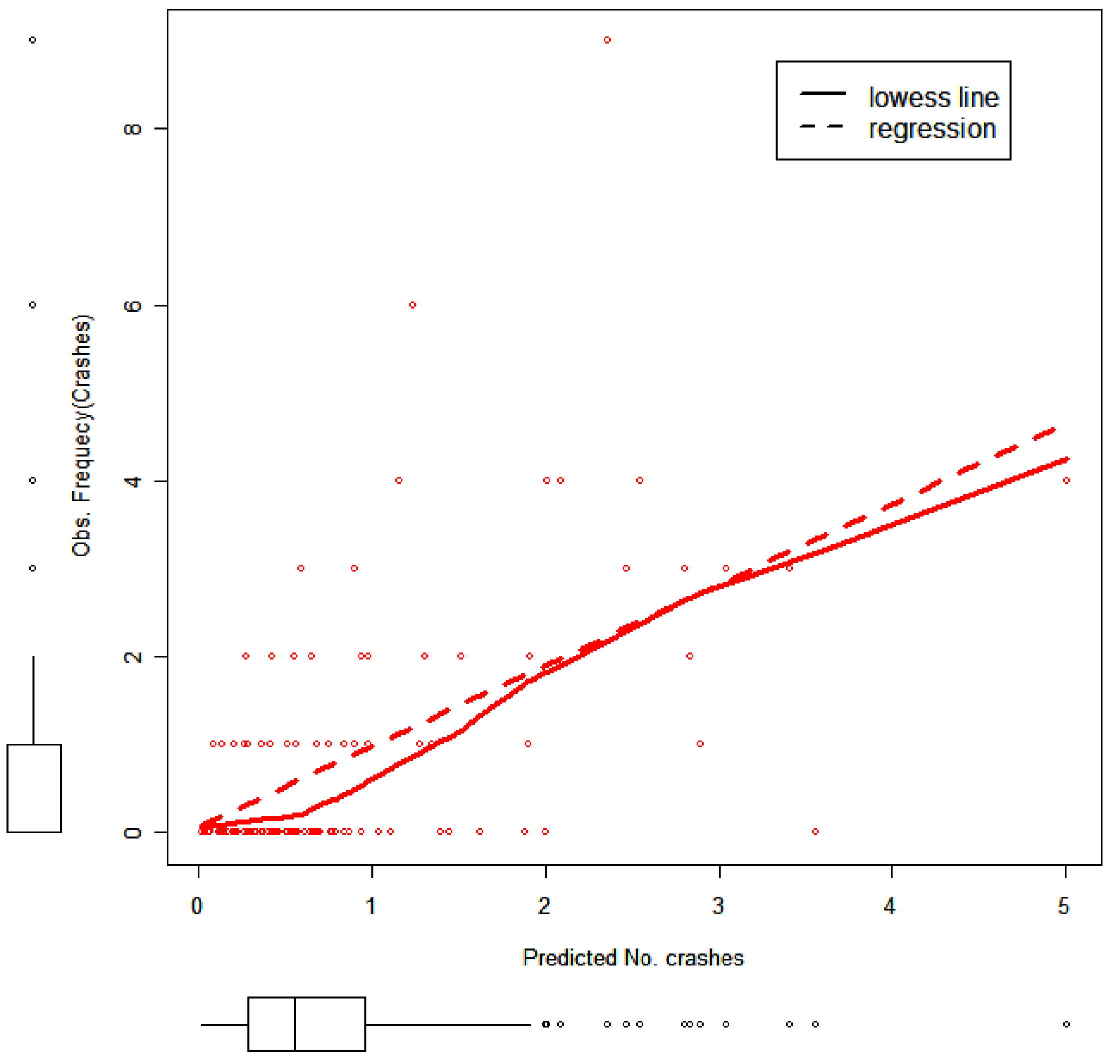

FIGURE 26 Model VII

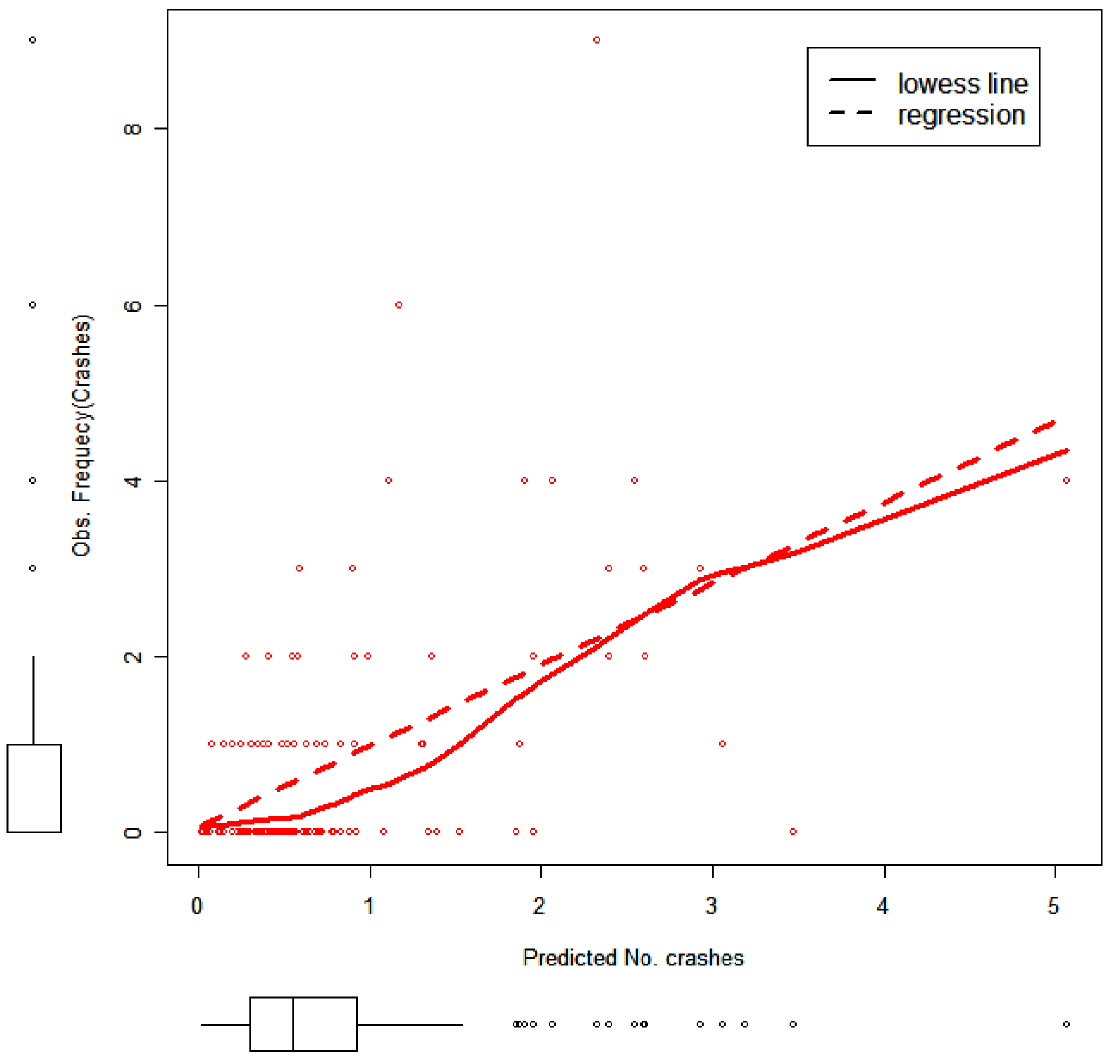

FIGURE 27 Model VIII

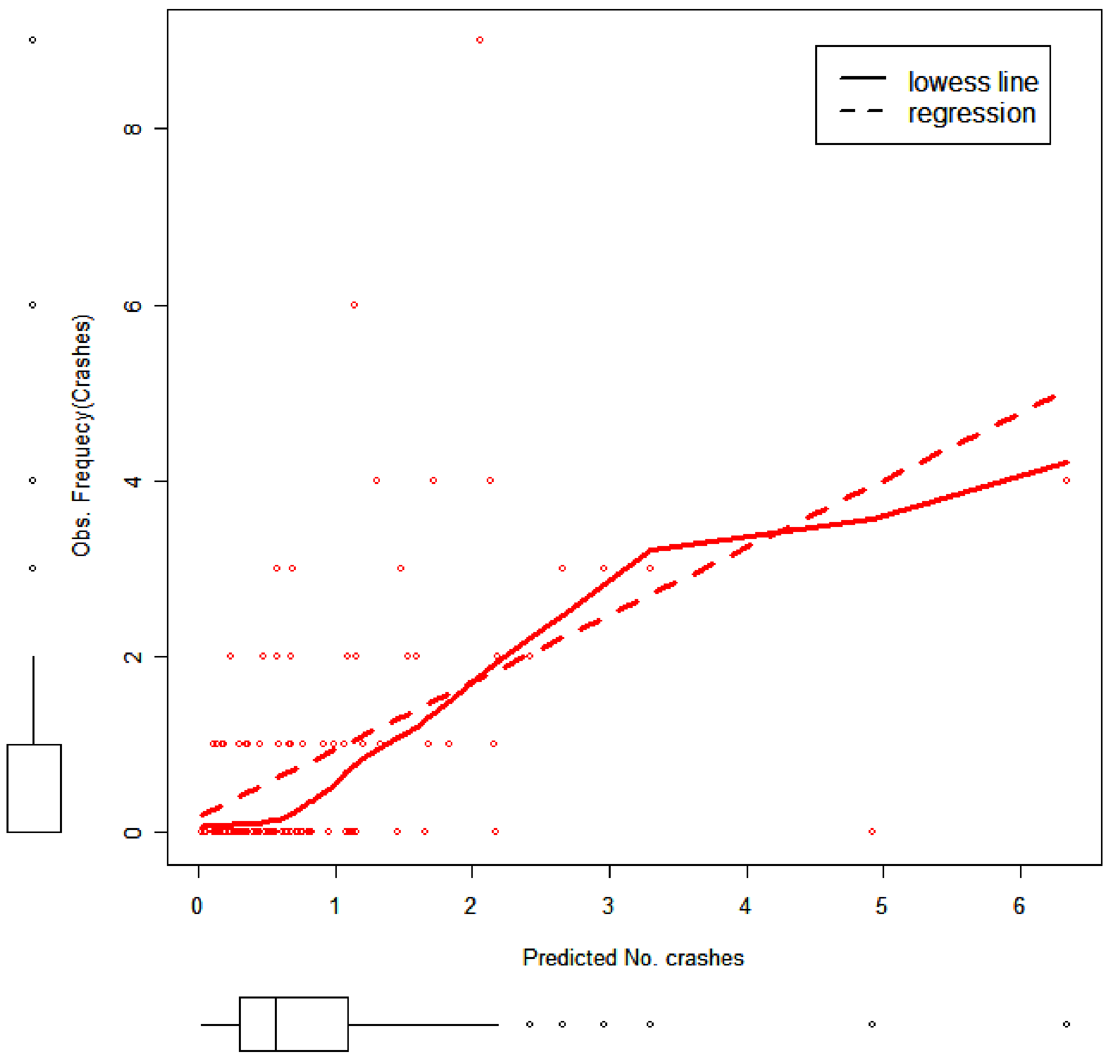

FIGURE 28 Model IX

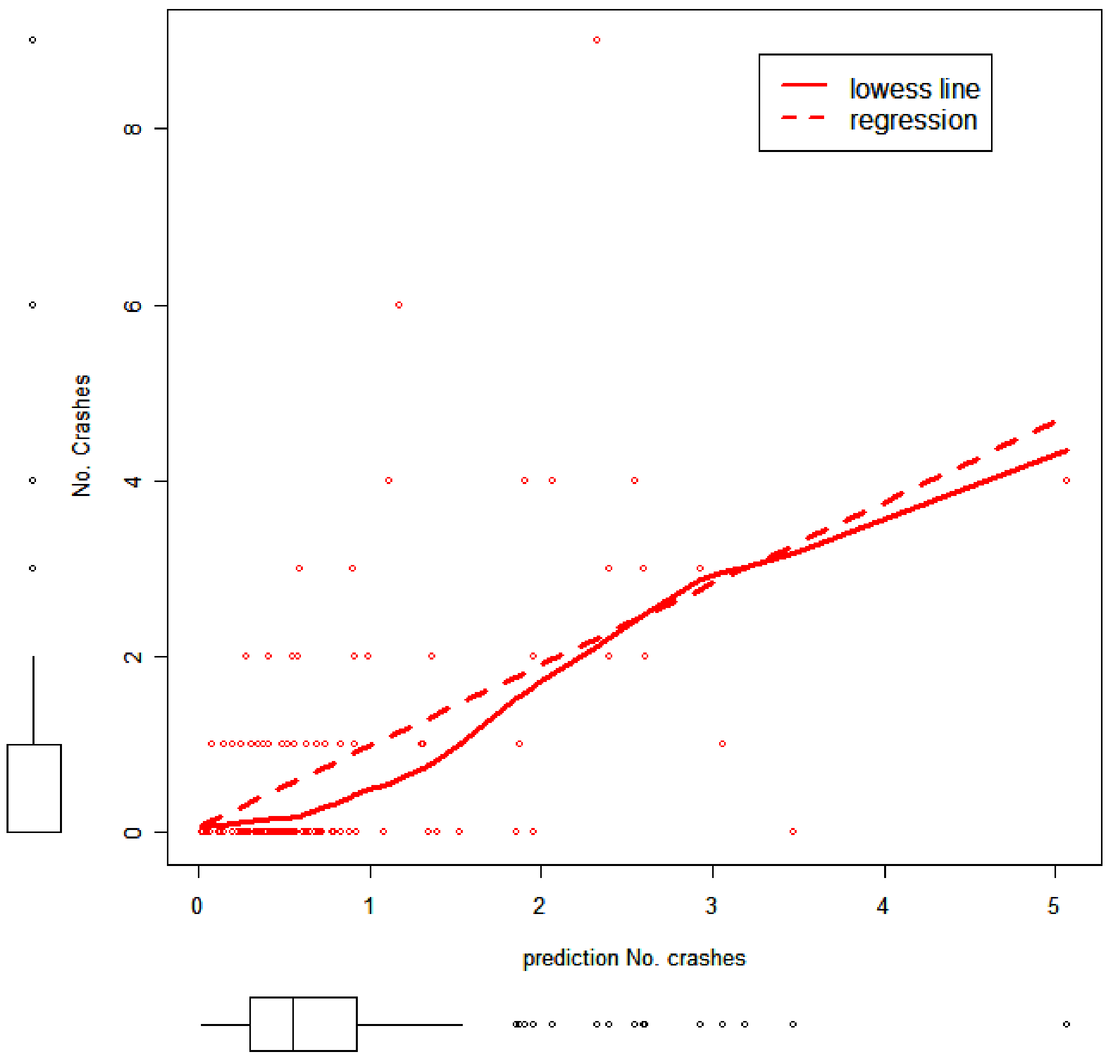

FIGURE 29 Model X

VITA

Name: Sunghoon Lee

Address: Department of Civil Engineering Texas A&M University 3136 TAMU

College Station, Texas 77843-3136

Email Address: valtope@neo.tamu.edu

Education: B.S., Urban Planning, Hong-ik University, Seoul Korea

M.S. Civil Engineering, Texas A&M University, 2007